BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of The original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
Please do not report the images to the
Image Problem Mailbox.**

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 August 2000 (24.08.2000)

PCT

(10) International Publication Number
WO 00/49145 A3

(51) International Patent Classification[7]: C12N 15/12, C07K 14/47, A61K 38/17, A01K 67/027, C07K 16/18, G01N 33/68, A61K 39/395, C12Q 1/68, A61K 49/00

(21) International Application Number: PCT/CA00/00171

(22) International Filing Date: 18 February 2000 (18.02.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/120,972 19 February 1999 (19.02.1999) US

(71) Applicant *(for all designated States except US)*: **MOUNT SINAI HOSPITAL** [CA/CA]; Samuel Lunenfeld Research Institute Office of Technology Transfer & Industrial Liaison, 600 University Avenue, Toronto, Ontario M5G 1X5 (CA).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **HIDAKA, Michihiro** [JP/JP]; The Second Department of Internal Medicine, Kumamoto University School of Medicine, Kumamoto, Kumamoto 860 (JP). **STANFORD, William** [US/CA]; 247 Wright Avenue, Toronto, Ontario M6R 1L4 (CA). **CARUANA, Georgina** [AU/CA]; 57 Charles St. West, Toronto, Ontario M5S 2X1 (CA). **KIMURA, Yuki** [JP/CA]; 45 Eastmount Avenue, Toronto, Ontario M4K 1V2 (CA).

(74) Agents: VAN ZANT, Joan, M. et al.; Swabey Ogilvy Renault, Suite 1600 - 1981 Mcgill College Avenue, Montreal, Quebec H3A 2Y3 (CA).

(81) Designated States *(national)*: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *With international search report.*

(88) Date of publication of the international search report:
4 January 2001

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 00/49145 A3

(54) Title: HEMATOPOIETIC GENES HZF AND HHL AND POLYPEPTIDES ENCODED THEREOF

(57) Abstract: The invention also relates to novel genes primarily expressed in hematopoietic lineages, polypeptides encoded by the novel genes and truncations, analogs, homologs, and isoforms of the polypeptides; and, uses of the polypeptides and genes.

INTERNATIONAL SEARCH REPORT

International Application No
PC CA 00/00171

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/12 C07K14/47 A61K38/17 A01K67/027 C07K16/18
G01N33/68 A61K39/395 C12Q1/68 A61K49/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C07K C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	W.L. STANFORD ET AL: "Expression Trapping: Identification of novel genes expressed in hematopoietic and endothelial lineages by gene Trapping in ES cells" BLOOD, vol. 92, no. 12, 15 December 1998 (1998-12-15), pages 4622-4631, XP002138431 cited in the application the whole document --- -/--	1,3-9, 11-30

[X] Further documents are listed in the continuation of box C.

[ ] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
25 May 2000	20.09.00

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	LE CORNEC N.D.R.

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	M. MARRA ET AL: "The WashU-HHMI mouse EST project" EMBL DATABASE ENTRY MMA30065, ACCESSION NUMBER A030065, 20 August 1996 (1996-08-20), XP002138432 "98% identity in 497 bp with sequence ID no.1" & UNPUBLISHED, ---	1,3,4,7, 8
A	R.K. BAKER ET AL: "In vitro preselection of gene-trapped embryonic Stem cell clones for characterizing novel developmentally regulated genes in the mouse" DEVELOPMENTAL BIOLOGY, vol. 185, no. 2, 15 May 1997 (1997-05-15), pages 201-214, XP000907315 cited in the application ---	
A	K. MUTH ET AL: "Disruption of genes regulated during hematopoietic differentiation of mouse embryonic stem cells" DEVELOPMENTAL DYNAMICS, vol. 212, 1998, pages 277-283, XP000911107 cited in the application ---	
A	SCOTT L M ET AL: "E3, A HEMATOPOIETIC-SPECIFIC TRANSCRIPT DIRECTLY REGULATED BY THE RETINOIC ACID RECEPTOR ALPHA" BLOOD,US,W.B. SAUNDERS, PHILADELPHIA, VA, vol. 88, no. 7, 1 October 1996 (1996-10-01), pages 2517-2530, XP002048663 ISSN: 0006-4971 ---	
A	TORU NAKANO ET AL: "generation of Lymphohematopoietic cells from embryonic stem cells in culture" SCIENCE., vol. 265, 19 August 1994 (1994-08-19), pages 1098-1101, XP002138433 AAAS. LANCASTER, PA., US cited in the application ---	
A	C. ZHANG ET AL: "activation of the megakaryocyte-specific gene platelet Basic protein (PBP) by the ets family factor PU.1" JOURNAL OF BIOLOGICAL CHEMISTRY., vol. 272, 1997, pages 26236-26246, XP002138434 AMERICAN SOCIETY OF BIOLOGICAL CHEMISTS, BALTIMORE, MD., US ISSN: 0021-9258 ---	

-/--

1

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	M. HIDAKA ET AL: "Gene trapping of two novel genes, Hzf and Hhl, expressed in hematopoietic cells" MECHANISMS OF DEVELOPMENT, vol. 90, no. 1, January 2000 (2000-01), pages 3-15, XP000911109 the whole document ---	1,3-9, 11-30
T	T. ERA ET AL: "Characterization of hematopoietic lineage-specific gene expression by ES cell in vitro differentiation induction system" BLOOD, vol. 95, no. 3, 1 February 2000 (2000-02-01), pages 870-878, XP002138436 -----	

INTERNATIONAL SEARCH REPORT

Im. .ational application No.
PCT/CA 00/00171

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

see FURTHER INFORMATION sheet PCT/ISA/210

2. [X] Claims Nos.: 23-26 all partially
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

see FURTHER INFORMATION sheet PCT/ISA/210

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [X] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Claims 1, 9, 29, 30 (all completely) and 3-8, 11-28 (all partially)

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.1

Although claims 23-24 are directed to a method of treatment of the human/animal body (rule 39.1 IV PCT), the search has been carried out and based on the alleged effects of the compound/composition.
Although claims 17 and 18 (as far as they concern an in vivo method) are directed to a diagnostic method practised on the human/animal body (rule 39.1 IV PCT), the search has been carried out and based on the alleged effects of the compound/composition.

Continuation of Box I.2

Claims Nos.: 23-26 all partially

Present claims 23-26 relate to a method or to a composition referring to a substance or compound identified using a method defined in claim 19 or 20, without giving a true technical characterization.
Moreover, no such substance or compound is defined in the application. In consequence, the scope of said claims is ambiguous and vague, and their subject-matter is not sufficiently disclosed and supported.
No search can be carried out for such purely speculative claims whose wording is, in fact, a mere recitation of the results to be achieved : A partial search has been carried out as far as the substance or compound relates to an antibody against Hzf. .

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

1. Claims: (1, 9, 29, 30) all completely and (3-8, 11-28) all partially

Hzf nucleic acid represented by sequence ID no.1 and polypeptide represented by sequence ID no.2. Vector, Host cell. Antibody. Therapeutic and diagnostic uses.

2. Claims: (2,10) all completely and (3-8,11-28) partially

Hhl nucleic acid represented by sequence ID no.3 and polypeptide represented by sequence ID no.4. Vector, Host cell. Antibody. Therapeutic and diagnostic uses.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 :
C12N 15/12, C07K 14/47, A61K 38/17, A01K 67/027, C07K 16/18, G01N 33/68, A61K 39/395, C12Q 1/68, A61K 49/00

A2

(11) International Publication Number: **WO 00/49145**

(43) International Publication Date: 24 August 2000 (24.08.00)

(21) International Application Number: PCT/CA00/00171

(22) International Filing Date: 18 February 2000 (18.02.00)

(30) Priority Data:
60/120,972 19 February 1999 (19.02.99) US

(71) Applicant *(for all designated States except US)*: MOUNT SINAI HOSPITAL [CA/CA]; Samuel Lunenfeld Research Institute Office of Technology Transfer & Industrial Liaison, 600 University Avenue, Toronto, Ontario M5G 1X5 (CA).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: HIDAKA, Michihiro [JP/JP]; The Second Department of Internal Medicine, Kumamoto University School of Medicine, Kumamoto, Kumamoto 860 (JP). STANFORD, William [US/CA]; 247 Wright Avenue, Toronto, Ontario M6R 1L4 (CA). CARUANA, Georgina [AU/CA]; 57 Charles St. West, Toronto, Ontario M5S 2X1 (CA). KIMURA, Yuki [JP/CA]; 45 Eastmount Avenue, Toronto, Ontario M4K 1V2 (CA).

(74) Agents: VAN ZANT, Joan, M. et al.; Swabey Ogilvy Renault, Suite 1600 - 1981 Mcgill College Avenue, Montreal, Quebec H3A 2Y3 (CA).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: NOVEL HEMATOPOIETIC GENES AND POLYPEPTIDES

(57) Abstract

The invention also relates to novel genes primarily expressed in hematopoietic lineages, polypeptides encoded by the novel genes and truncations, analogs, homologs, and isoforms of the polypeptides; and, uses of the polypeptides and genes.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

Title: Novel Hematopoietic Genes and Polypeptides

FIELD OF THE INVENTION

The invention relates to novel genes primarily expressed in hematopoietic cells, polypeptides encoded by the novel genes and truncations, analogs, homologs, and isoforms of the polypeptides; and, uses of the polypeptides and genes.

BACKGROUND OF THE INVENTION

Many of the molecular events involved in the embryonic development and differentiation of hematopoietic cells remain to be elucidated. In the mouse, several genetic approaches have been employed in an attempt to identify and determine the function of novel hematopoietic genes. One approach involves the cloning of mutant loci responsible for altered hematopoietic phenotypes in mice harboring naturally occurring or induced mutations [1,2,3]. A complementary approach involves the generation of new mutations in the mouse germline by homologous recombination in embryonic stem (ES) cells to mutate candidate hematopoietic genes [4,5,6,7,8,9,10]. Both approaches are labour intensive, the former being restricted to the repetoire of mutant phenotypes while the latter is restricted to previously identified genes. Thus, neither approach is easily applied on a genome-wide level. In contrast, gene trapping in ES cells provides an effective approach to screen and analyze novel genes because it simultaneously provides information on the pattern of expression *in vivo*, sequence and resulting mouse mutant phenotype [11,12,13,14,15,16,17,18].

One potential drawback of gene trapping is the very large size of the mammalian genome and hence the integration of the gene trap vector into genes that may not be of immediate interest. To pre-select for genes that are preferentially expressed in hematopoietic cells, an expression trapping strategy has been developed that takes advantage of the developmental potential of ES cells to differentiate into diverse cell lineages *in vitro*, including hematopoietic cells [19,20,21,22,23,24,25]. Upon the removal of leukemia inhibitory factor (LIF), ES cells spontaneously differentiate into embryoid bodies (EBs), structures which resemble postimplantation embryos, comprising a number of cell lineages and structures, including blood islands which contain hematopoietic precursor cells. Secondary plating of cells obtained from dissociated EBs in semisolid cultures, containing various cytokines and growth factors, results in the growth of colonies of erythroid, myeloid and lymphoid lineages [19,24,25,26,27,28]. Recently, *in vitro* differentiation of ES cells has been combined with gene trapping strategies to pre-screen for novel genes which respond to exogenous factors [29,17] or are regulated in a tissue-specific manner prior to transmission into the mouse germline [18, 30,31,32]. In order to screen large numbers of cell clones based on the expression of trapped genes in hematopoietic and endothelial cells, an expression trap screen has been combined with an attached EB culture system to identify genes expressed within blood islands and endothelial cells [18]. Muth *et al.*,(1998) [32] have developed a related approach to identify genes that are repressed during hematopoietic differentiation.

SUMMARY OF THE INVENTION

The ES cell-OP9 co-culture system has been combined with gene trapping to screen for genes involved in the development and differentiation of hematopoietic cells. The OP9 stromal cell line induces

ES cell differentiation into mesodermal colonies that, when replated, differentiate into single lineage precursors, without the addition of exogenous growth factors [33]. Using this system, the present inventors have identified two novel genes designated *Hzf* (hematopoietic zinc finger) and *Hhl* (denoting the embryonic expression seen within hematopoietic cells, heart and liver) which are expressed in a regulated pattern during hematopoietic differentiation *in vitro*. These genes exhibit hematopoietic-specific expression *in vivo*, indicating that the *in vitro* pre-screening strategy successfully predicts the expression of trapped genes *in vivo*. Within the hematopoietic compartment, both genes were predominantly expressed in megakaryocytes.

Broadly stated the present invention contemplates an isolated hematopoietic nucleic acid molecule encoding a Hzf or Hhl polypeptide of the invention, including mRNAs, DNAs, cDNAs, genomic DNAs, PNAs, as well as antisense analogs and biologically, diagnostically, prophylactically, clinically or therapeutically useful variants or fragments thereof, and compositions comprising same.

The invention also contemplates isolated hematopoietic Hzf or Hzl polypeptides encoded by a nucleic acid molecule of the invention a truncation, an analog, an allelic or species variation thereof, or a homolog of a polypeptide of the invention or a truncation thereof. (Truncations, analogs, allelic or species variations, and homologs are collectively referred to herein as "Hzf Related Polypeptides" or "Hhl Related Polypeptides).

The nucleic acid molecules of the invention permit identification of untranslated nucleic acid sequences or regulatory sequences that specifically promote expression of genes operatively linked to the promoter regions. Identification and use of such promoter sequences are particularly desirable in instances, such as gene transfer or gene therapy, which may specifically require heterologous gene expression in a limited environment e.g hematopoietic system. The invention therefore contemplates a nucleic acid molecule comprising a non-coding sequence such as a 5' and/or 3" sequence, preferably a non-coding sequence of *Hzf or Hhl.*

The nucleic acid molecules which encode for the mature Hzf or Hhl polypeptide (may include only the coding sequence for the mature polypeptide; the coding sequence for the mature polypeptide and additional coding sequences (e.g. leader or secretory sequences, propolypeptide sequences); the coding sequence for the mature polypeptide (and optionally additional coding sequence) and non-coding sequence, such as introns or non-coding sequence 5' and/or 3' of the coding sequence of the mature polypeptide.

Therefore, the term "nucleic acid molecule encoding a polypeptide" encompasses a nucleic acid molecule which includes only coding sequence for the polypeptide as well as a nucleic acid molecule which includes additional coding and/or non-coding sequences.

The nucleic acid molecules of the invention may be inserted into an appropriate vector, and the vector may contain the necessary elements for the transcription and translation of an inserted coding sequence. Accordingly, vectors may be constructed which comprise a nucleic acid molecule of the invention, and where appropriate one or more transcription and translation elements linked to the nucleic acid molecule.

Vectors are contemplated within the scope of the invention which comprise regulatory sequences of the invention, as well as chimeric gene constructs wherein a regulatory sequence of the invention is operably linked to a heterologous nucleic acid, and a transcription termination signal.

A vector can be used to transform host cells to express a Hzf or Hhl Polypeptide or a Hzf or Hhl Related Polypeptide, or a heterologous polypeptide (i.e. a polypeptide not naturally in the host cell). Therefore, the invention further provides host cells containing a vector of the invention. The invention also contemplates transgenic non-human mammals whose germ cells and somatic cells contain a vector comprising a nucleic acid molecule of the invention in particular one that encodes an analog of Hzf or Hhl, or a truncation of Hzf or Hhl.

The polypeptides of the invention may be obtained as an isolate from natural cell sources, but they are preferably produced by recombinant procedures. In one aspect the invention provides a method for preparing a Hzf or Hhl Polypeptide, or a Hzf or Hhl Related Polypeptide utilizing the purified and isolated nucleic acid molecules of the invention. In an embodiment a method for preparing a Hzf or Hhl Polypeptide, or a Hzf or Hhl Related Polypeptide is provided comprising:

(a) transferring a vector of the invention comprising a nucleic acid sequence encoding a Hzf or Hhl Polypeptide, or a Hzf or Hhl Related Polypeptide, into a host cell;

(b) selecting transformed host cells from untransformed host cells;

(c) culturing a selected transformed host cell under conditions which allow expression of the Hzf or Hhl Polypeptide, or a Hzf or Hhl Related Polypeptide; and

(d) isolating the Hzf or Hhl Polypeptide, or Hzf or Hhl Related Polypeptide.

The invention further broadly contemplates a recombinant Hzf or Hhl Polypeptide, or Hzf or Hhl Related Polypeptide obtained using a method of the invention.

A Hzf or Hhl Polypeptide, or Hzf or Hhl Related Polypeptide of the invention may be conjugated with other molecules, such as polypeptides, to prepare fusion polypeptides or chimeric polypeptides. This may be accomplished, for example, by the synthesis of N-terminal or C-terminal fusion polypeptides.

The invention further contemplates antibodies having specificity against an epitope of a Hzf or Hhl Polypeptide, or a Hzf or Hhl Related Polypeptide of the invention. Antibodies may be labeled with a detectable substance and used to detect polypeptides of the invention in biological samples, tissues, and cells.

The invention also permits the construction of nucleotide probes that are unique to nucleic acid molecules of the invention and/or to polypeptides of the invention. Therefore, the invention also relates to a probe comprising a sequence encoding a polypeptide of the invention, or a portion (i.e. fragment) thereof. The probe may be labeled, for example, with a detectable substance and it may be used to select from a mixture of nucleic acid molecules a nucleic acid molecule of the invention including nucleic acid molecules coding for a polypeptide which displays one or more of the properties of a polypeptide of the invention.

In accordance with an aspect of the invention there is provided a method of, and products for diagnosing and monitoring conditions mediated by Hzf or Hhl by determining the presence of nucleic acid molecules and polypeptides of the invention.

Still further the invention provides a method for evaluating a test substance or compound for its ability to modulate the biological activity of a Hzf or Hhl Polypeptide, or a Hzf or Hhl Related Polypeptide of the invention. For example, a substance or compound which inhibits or enhances the activity of a Hzf or Hhl Polypeptide, or a Hzf or Hhl Related Polypeptide may be evaluated. "Modulate" refers to a change or an alteration in the biological activity of a polypeptide of the invention. Modulation may be an increase or a decrease in activity, a change in characteristics, or any other change in the biological, functional, or immunological properties of the polypeptide.

Compounds which modulate the biological activity of a polypeptide of the invention may also be identified using the methods of the invention by comparing the pattern and level of expression of a nucleic acid molecule or polypeptide of the invention in biological samples, tissues and cells, in the presence, and in the absence of the compounds.

Methods are also contemplated that identify compounds or substances (e.g. polypeptides) which interact with Hzf or Hhl regulatory sequences (e.g. promoter sequences, enhancer sequences, negative modulator sequences).

The nucleic acid molecules, polypeptides, and substances and compounds identified using the methods of the invention, may be used to modulate the biological activity of a Hzf or Hhl Polypeptide, or a Hzf or Hhl Related Polypeptide of the invention, and they may be used in the treatment of conditions mediated by Hzf or Hhl such as hematopoietic disorders. Accordingly, the nucleic acid molecules, polypeptides, substances and compounds may be formulated into compositions for administration to individuals suffering from one or more of these conditions. Therefore, the present invention also relates to a composition comprising one or more of a polypeptide, nucleic acid molecule, or substance or compound identified using the methods of the invention, and a pharmaceutically acceptable carrier, excipient or diluent. A method for treating or preventing these conditions is also provided comprising administering to a patient in need thereof, a composition of the invention.

The present invention in another aspect provides means necessary for production of gene-based therapies directed at the hematopoietic system. These therapeutic agents may take the form of polynucleotides comprising all or a portion of a nucleic acid molecule of the invention comprising a regulatory sequence of Hzf or Hhl placed in appropriate vectors or delivered to target cells in more direct ways.

In accordance with a further aspect of the invention, there are provided processes for utilizing polypeptides or nucleic acid molecules, for *in vitro* purposes related to scientific research, synthesis of DNA and manufacture of vectors.

These and other aspects, features, and advantages of the present invention should be apparent to those skilled in the art from the following drawings and detailed description.

DESCRIPTION OF THE DRAWINGS

The invention will be better understood with reference to the drawings in which:

Figure 1. *LacZ* Expression Pattern of Clone Hhl (A-D) and Hzf (E-F) during *in vitro* Hematopoietic Differentiation on OP9 stromal cells. β-galactosidase (β-gal) activity within the gene trap

clone Hhl was detected around day 5 within mesodermal colonies and was maintained in hematopoietic cells (day 12). The gene trap clone Hzf expressed low levels of β-gal activity in undifferentiated ES cells which was maintained within hematopoietic cells (day 2-12). Cultures were fixed with 0.25% gluteraldehyde and stained with X-gal after 2 days (A, E), 5 days (B, F), 8 days (C, G) and 12 days (D, H). Bar =50mm.

Figure 2. *LacZ* Expression in Hhl (A-D) and Hzf (E-H) Heterozygous Embryos.
Wholemount *lacZ* expression within 9.5 d.p.c (A) Hhl heterozygous F_1 embryos demonstrated expression within the yolk sac and heart and also within the liver at 14.5 d.p.c (B). Histological analysis of the liver (C) and heart (D) of 14.5 d.p.c Hhl embryos. Wholemount *lacZ* expression within 9.5 d.p.c Hhl F_1 heterozygous embryos (E) demonstrated expression within the somites, basal ganglia, apical ectodermal region of the limb buds and liver primordium and in the skin (Figure 2H). At 14.5 d.p.c (F) expression was seen in the trigeminal ganglia, thymus, salivary gland, spinal cord and spotty staining within the fetal liver. Histological analysis of the liver (G) and skin (H) of 14.5d.p.c Hzf embryo. Bar=100mm (A,C,E,H), 1mm (B, F), 500mm (D), 50mm (G).

Figure 3. *LacZ* Expression within Bone Marrow Cells derived from Hhl and Hzf Heterozygous Adult Mice. X-gal staining of Hhl (A) and Hzf (B) bone marrow cells. *LacZ* expression detected by FDG staining (C) and cell morphology analysed by Wright-Giemsa staining (D) of the same sections of bone marrow cells from Hzf mice demonstrating the coincidence of *lacZ* expression within cells morphologically resembling megakaryocytes. Hhl bone marrow sections also demonstrated the same coincidence of expression within megakaryocytes (not shown). Bar = 50mm.

Figure 4. *LacZ* Expression within Bone Marrow Cell derived Colonies. Bone marrow cells from Hzf and Hhl adult heterozygous mice were cultured in semi-solid agarose in the presence of TPO, IL-3 and SLF for the analysis of CFU-Mk. For the analysis of BFU-E, CFU-G, M, GM and GEMM bone marrow cells were plated in methylcellulose containing IL-3, IL-6, Epo and SLF. Colonies were either X-gal stained directly in the cultures or picked and then stained. High levels of X-gal staining were seen in CFU-Mk and CFU-GEMM for both Hzf and Hhl derived colonies. X-gal staining was seen in Hhl derived BFU-E but not in those derived from Hzf bone marrow cells. A low frequency of X-gal staining was detected in the macrophage component (depicted by the arrows) of CFU-M and GM for Hhl and only in CFU-M for Hzf. Photographs of cells from representative colonies are shown.

Figure 5. 5'-RACE Sequence of *Hhl* and *Hzf*.
(5A) Schematic diagram of the PT1/ATG gene trap vector and 5'-RACE methodology.
(5B) The PT1/ATG vector contains a promoterless *lacZ* gene immediately upstream of a splice acceptor (SA) site and the neo^R gene driven by the PGK-1 promoter. Random integration within an intron will generate a spliced fusion transcript between *lacZ* and the endogenous gene located at the site of vector integration. The fusion with *lacZ* enables primers to be designed (a, b and e) to clone part of the upstream fused exon of the trapped gene by 5' rapid amplification of cDNA ends (RACE). The Kpn1 and Pst 1 restriction sites and primers (c & d, GTlacz-3 and 4, respectively) used in the inverse PCR strategy for *Hhl* are also indicated.

(5C) 5'- RACE sequence of *Hhl* (bold). Arrows indicate sequences of the 12G10-5 and 12G10-6 primers used for 3'-RACE. (C) 5'- RACE sequence of *Hzf* (bold). Arrows indicate sequences of 4H5-1 and 4H5-2 primers used for 3'-RACE

Figure 6. Nucleotide Sequence and the Deduced Amino Acid Sequence of the *Hhl* and *Hzf* genes. (6A) The putative initiation codon of *Hhl* is located at bases 266-268. The *Hhl* open reading frame (ORF) encodes 298 amino acids. (6B) The longest ORF of *Hzf* begins with the ATG at nucleotide 29 and encodes 396 amino acids. The polyadenylation signal is underlined and the three C_2H_2-type zinc finger motifs are boxed. The arrows depict the site of vector integration.

Figure 7. Northern blot analysis of total RNA from wild type mouse adult tissues using probes from the *Hhl* (A) and *Hzf* (B) genes. The *GAPDH* probe was used as an internal loading control (C).

Figure 8. RT-PCR Analysis of Hzf and Hhl Expression in Normal and Trapped Mice. RNA was isolated from brain and amplified by RT-PCR from genotyped Hzf and Hhl mice. Amplified products were then southern blotted and hybridized with gene specific or *lacZ* probes. Amplified products were obtained from Hhl (A) and Hzf (B) homozygous mice using a primer 5' and another 3' to the vector integration site demonstrating that endogenous transcript 3' to the integration site had not been disrupted. Wild-type mice were distinguished from mutant mice due to the lack of amplification of a *Hzf* or *Hhl-lacZ* fusion product using a primer 5' to the integration site and another complementary to the *lacZ* reporter gene.

DETAILED DESCRIPTION OF THE INVENTION

In accordance with the present invention there may be employed conventional molecular biology, microbiology, and recombinant DNA techniques within the skill of the art. Such techniques are explained fully in the literature. See for example, Sambrook, Fritsch, & Maniatis, Molecular Cloning: A Laboratory Manual, Second Edition (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y); DNA Cloning: A Practical Approach, Volumes I and II (D.N. Glover ed. 1985); Oligonucleotide Synthesis (M..J. Gait ed. 1984); Nucleic Acid Hybridization B.D. Hames & S.J. Higgins eds. (1985); Transcription and Translation B.D. Hames & S.J. Higgins eds (1984); Animal Cell Culture R.I. Freshney, ed. (1986); Immobilized Cells and enzymes IRL Press, (1986); and B. Perbal, A Practical Guide to Molecular Cloning (1984).

Nucleic Acid Molecules and Polypeptides of the Invention

Nucleic Acid Molecules

As hereinbefore mentioned, the invention provides isolated Hzf and Hhl nucleic acid molecules. The term "isolated" refers to a nucleic acid (or polypeptide) removed from its natural environment, purified or separated, or substantially free of cellular material or culture medium when produced by recombinant DNA techniques, or chemical reactants, or other chemicals when chemically synthesized. Preferably, an isolated nucleic acid is at least 60% free, more preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated. The term "nucleic acid" is intended to include modified or unmodified DNA, RNA, including mRNAs, DNAs, cDNAs, and genomic DNAs, or a mixed polymer, and can be either single-stranded, double-stranded or triple-stranded. For example, a nucleic acid sequence may be a single-stranded or double-stranded DNA, DNA that is a mixture

of single-and double-stranded regions, or single-, double- and triple-stranded regions, single- and double-stranded RNA, RNA that may be single-stranded, or more typically, double-stranded, or triple-stranded, or a mixture of regions comprising RNA or DNA, or both RNA and DNA. The strands in such regions may be from the same molecule or from different molecules. The DNAs or RNAs may contain one or more modified bases. For example, the DNAs or RNAs may have backbones modified for stability or for other reasons. A nucleic acid sequence includes an oligonucleotide, nucleotide, or polynucleotide. The term "nucleic acid molecule" and in particular DNA or RNA, refers only to the primary and secondary structure and it does not limit it to any particular tertiary forms.

In an embodiment of the invention an isolated nucleic acid molecule is contemplated which comprises:

(i) a nucleic acid sequence encoding a polypeptide having substantial sequence identity with the amino acid sequence of SEQ. ID. NO. 2, or 4;

(ii) a nucleic acid sequence complementary to (i);

(iii) a nucleic acid sequence differing from any of (i) or (ii) in codon sequences due to the degeneracy of the genetic code;

(iv) a nucleic acid sequence comprising at least 10, 15, 18, and preferably at least 20 nucleotides capable of hybridizing to a nucleic acid sequence of SEQ. ID. NO. 1 or 3 or to a degenerate form thereof;

(v) a nucleic acid sequence encoding a truncation, an analog, an allelic or species variation of a polypeptide comprising the amino acid sequence of SEQ. ID. NO. 2 or 4; or

(vi) a fragment, or allelic or species variation of (i), (ii) or (iii)

In a specific embodiment, the isolated nucleic acid molecule comprises:

(i) a nucleic acid sequence having substantial sequence identity or sequence similarity with a nucleic acid sequence of SEQ. ID. NO. 1 or 3;

(ii) nucleic acid sequences complementary to (i), preferably complementary to the full nucleic acid sequence of SEQ. ID. NO. 1 or 3;

(iii) nucleic acid sequences differing from any of the nucleic acid sequences of (i) or (ii) in codon sequences due to the degeneracy of the genetic code; or

(iv) a fragment, or allelic or species variation of (i), (ii) or (iii).

The term "complementary" refers to the natural binding of nucleic acid molecules under permissive salt and temperature conditions by base-pairing. For example, the sequence "A-G-T" binds to the complementary sequence "T-C-A" .Complementarity between two single-stranded molecules may be "partial", in which only some of the nucleic acids bind, or it may be complete when total complementarity exists between the single stranded molecules.

In a preferred embodiment the isolated nucleic acid comprises a nucleic acid sequence encoded by the amino acid sequence of SEQ. ID. NO. 2 or 4, or comprises the nucleic acid sequence of SEQ. ID. NO. 1 or 3 wherein T can also be U.

The terms "sequence similarity" or "sequence identity" refer to the relationship between two or more amino acid or nucleic acid sequences, determined by comparing the sequences, which relationship is generally known as "homology". Identity in the art also means the degree of sequence relatedness between amino acid or nucleic acid sequences, as the case may be, as determined by the match between strings of such sequences. Both identity and similarity can be readily calculated (Computational Molecular Biology, Lesk, A.M., ed., Oxford University Press New York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D.W. ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part I, Griffin, A.M., and Griffin, H.G. eds. Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, New York, 1987; and Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds. M. Stockton Press, New York, 1991). While there are a number of existing methods to measure identity and similarity between two amino acid sequences or two nucleic acid sequences, both terms are well known to the skilled artisan (Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, New York, 1987; Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds. M. Stockton Press, New York, 1991; and Carillo, H., and Lipman, D. SIAM J. Applied Math., 48:1073, 1988). Preferred methods for determining identity are designed to give the largest match between the sequences tested. Methods to determine identity are codified in computer programs. Preferred computer program methods for determining identity and similarity between two sequences include but are not limited to the GCG program package (Devereux, J. et al, Nucleic Acids Research 12(1): 387, 1984), BLASTP, BLASTN, and FASTA (Atschul, S.F. et al., J. Molec. Biol. 215:403, 1990). Identity or similarity may also be determined using the alignment algorithm of Dayhoff et al [Methods in Enzymology 91: 524-545 (1983)].

Preferably, the nucleic acids of the present invention have substantial sequence identity using the preferred computer programs cited herein, for example greater than 50% , 60%, 70%, 80%, or 90% sequence identity, and preferably at least 90% , 95%, 96%, 97%, 98%, or 99% sequence identity to the sequence shown in SEQ. ID. NO. 1 or 3.

Isolated nucleic acids encoding a Hzf or Hhl Polypeptide and comprising a sequence that differs from the nucleic acid sequence shown in SEQ. ID. NO. 1 or 3 due to degeneracy in the genetic code are also within the scope of the invention. Such nucleic acids encode equivalent polypeptides but differ in sequence from the sequence of SEQ. ID. NO. 1 or 3 due to degeneracy in the genetic code. As one example, DNA sequence polymorphisms within *hzf* or *hhl* may result in silent mutations that do not affect the amino acid sequence. Variations in one or more nucleotides may exist among individuals within a population due to natural allelic variation. Any and all such nucleic acid variations are within the scope of the invention. DNA sequence polymorphisms may also occur which lead to changes in the amino acid sequence of Hzf or Hhl Polypeptide. These amino acid polymorphisms are also within the scope of the present invention. In addition, species variations i.e. variations in nucleotide sequence naturally occurring among different species, are within the scope of the invention.

Another aspect of the invention provides a nucleic acid molecule which hybridizes under selective conditions, (e.g. high stringency conditions), to a nucleic acid which comprises a sequence which encodes

a Hzf or Hhl Polypeptide of the invention. Preferably the sequence encodes the amino acid sequence shown in SEQ. ID. NO. 2 or 4 and comprises at least 10, preferably at least 15, more preferably at least 18, and most preferably at least 20 nucleotides. Selectivity of hybridization occurs with a certain degree of specificity rather than being random. Appropriate stringency conditions which promote DNA hybridization are known to those skilled in the art, or can be found in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. For example, hybridization may occur at 30°C in 750 mM NaCl, 75mM trisodium citrate, and 1% SDS, preferably 37°C in 500mM NaCl, 500 mM trisodium citrate, 1% SDS, 35% formamide, and 100μg/ml denatured salmon sperm DNA (ssDNA), and more preferably 42°C in 250 mM NaCl, 25 mM trisodium citrate, 1% SDS, 50% formamide, and 200 μg/ml ssDNA. Useful variations on these conditions will be readily apparent to those skilled in the art.

The stringency may be selected based on the conditions used in the wash step. Wash step stringency conditions may be defined by salt concentration and by temperature. Generally, wash stringency can be increased by decreasing salt concentration or by increasing termperature. By way of example, a stringent salt concentration for the wash step is preferably less than about 30 mM NaCl and 3mM trisodium citrate, and more preferably less than about 15 mM NaCl and 1.5 mM trisodium citrate. Stringent temperature conditions will generally include temperatures of a least about 25°C, more preferably at least about 68°C. In a preferred embodiment, the wash steps will be carried out at 42°C in 15 mM NaCl, 1.5mM trisodium citrate, and 0.1% SDS. In a more preferred embodiment the wash steps are carried out at 68°C in 15 mM NaCl, 1.5mM trisodium citrate, and 0.1% SDS. Variations on these conditions will be readily apparent to those skilled in the art.

It will be appreciated that the invention includes nucleic acid molecules encoding a Hzf or Hhl Polypeptide, or a Hzf or Hhl Related Polypeptide, including truncations of the polypeptides, allelic and species variants, and analogs of the polypeptides as described herein. In particular, fragments of a nucleic acid of the invention are contemplated that are a stretch of at least about 10, 15, or 18, and most preferably at least 20 nucleotides, more typically at least 50 to 200 nucleotides but less than 2 kb. It will further be appreciated that variant forms of the nucleic acid molecules of the invention which arise by alternative splicing of an mRNA corresponding to a cDNA of the invention are encompassed by the invention.

An isolated nucleic acid molecule of the invention which comprises DNA can be isolated by preparing a labeled nucleic acid probe based on all or part of the nucleic acid sequence shown in SEQ. ID. NO. 1 or 3. The labeled nucleic acid probe is used to screen an appropriate DNA library (e.g. a cDNA or genomic DNA library). For example, a cDNA library can be used to isolate a cDNA encoding a Hzf or Hhl Polypeptide, or a Hzf or Hhl Related Polypeptide by screening the library with the labeled probe using standard techniques. Alternatively, a genomic DNA library can be similarly screened to isolate a genomic clone encompassing a *hzf* or *hhl* gene. Nucleic acids isolated by screening of a cDNA or genomic DNA library can be sequenced by standard techniques.

An isolated nucleic acid molecule of the invention that is DNA can also be isolated by selectively amplifying a nucleic acid of the invention. "Amplifying" or "amplification " refers to the production of

additional copies of a nucleic acid sequence and is generally carried out using polymerase chain reaction (PCR) technologies well known in the art (Dieffenbach, C. W. and G. S. Dveksler (1995) PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview, N.Y.). In particular, it is possible to design synthetic oligonucleotide primers from the nucleotide sequence shown in SEQ. ID. NO. 1 or (e.g. SEQ. ID. Nos. 5-14) for use in PCR. A nucleic acid can be amplified from cDNA or genomic DNA using these oligonucleotide primers and standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. cDNA may be prepared from mRNA, by isolating total cellular mRNA by a variety of techniques, for example, by using the guanidinium-thiocyanate extraction procedure of Chirgwin et al., Biochemistry, 18, 5294-5299 (1979). cDNA is then synthesized from the mRNA using reverse transcriptase (for example, Moloney MLV reverse transcriptase available from Gibco/BRL, Bethesda, MD, or AMV reverse transcriptase available from Seikagaku America, Inc., St. Petersburg, FL).

An isolated nucleic acid molecule of the invention which is RNA can be isolated by cloning a cDNA encoding a Hzf or Hhl Polypeptide, or a Hzf or Hhl Related Polypeptide into an appropriate vector which allows for transcription of the cDNA to produce an RNA molecule which encodes a Hzf or Hhl Polypeptide, or a Hzf or Hhl Related Polypeptide. For example, a cDNA can be cloned downstream of a bacteriophage promoter, (e.g. a T7 promoter) in a vector, cDNA can be transcribed *in vitro* with T7 polymerase, and the resultant RNA can be isolated by conventional techniques.

A nucleic acid molecule of the invention may be engineered using methods known in the art to alter the Hzf or Hhl encoding sequence for a variety of purposes including modification of the cloning, processing, and/or expression of the gene product. Procedures such as DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleic acid molecules. Mutations may be introduced by oligonucleotide-mediated site-directed mutagenesis to create for example new restriction sites, alter glycosylation patterns, change codon preference, or produce splice variants.

Nucleic acid molecules of the invention may be chemically synthesized using standard techniques. Methods of chemically synthesizing polydeoxynucleotides are known, including but not limited to solid-phase synthesis which, like peptide synthesis, has been fully automated in commercially available DNA synthesizers (See e.g., Itakura et al. U.S. Patent No. 4,598,049; Caruthers et al. U.S. Patent No. 4,458,066; and Itakura U.S. Patent Nos. 4,401,796 and 4,373,071).

Determination of whether a particular nucleic acid molecule is a *hzf or hhl* gene or encodes a Hzf or Hhl Polypeptide, or a Hzf or Hhl Related Polypeptide can be accomplished by expressing the cDNA in an appropriate host cell by standard techniques, and testing the expressed polypeptide in the methods described herein. A *Hzf or Hhl* cDNA or cDNA encoding a Hzf or Hhl Polypeptide, or a Hzf or Hhl Related Polypeptide can be sequenced by standard techniques, such as dideoxynucleotide chain termination or Maxam-Gilbert chemical sequencing, to determine the nucleic acid sequence and the predicted amino acid sequence of the encoded polypeptide.

The nucleic acid molecules of the invention may be extended using a partial nucleotide sequence and various PCR-based methods known in the art to detect uptstream sequences such as promoters and regulatory elements. For example, restriction-site PCR which uses universal and nested primers to amplify unknown sequences from genomic DNA within a cloning vector may be employed (See Sarkar, G, PCR Methods Applic. 2:318-322, 1993). Inverse PCR which uses primers that extend in divergent directions to amplify unknown sequences from a circularized template may also be used. The template in inverse PCR is derived from restriction fragments adjacent to known sequences in human and yeast artificial chromosome DNA (See e.g. Lagerstrom, M., at al, PCR Methods Applic. 1:111-119, 1991). Other methods for retrieving unknown sequences are known in the art (e.g. Parker, J.D. et al, Nucleic Acids Res. 19:305-306, 1991). In addition, PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto, California) may be used to walk genomic DNA. The method is useful in finding intron/exon junctions and avoids the need to screen libraries.

It is preferable when screening for full-length cDNAs to use libraries that have been size-selected to include larger cDNAs. For situations in which an oligo d(T) library does not yield a full-length cDNA, it is preferable to use random-primed libraries which often include sequences containing the 5' regions of genes. Genomic libraries may be useful for extending the sequence into 5'non-translated regulatory regions.

Commercially available capillary electrophoresis systems may be employed to analyse the size or confirm the sequence of PCR or sequencing products. The system may use flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Commercially available software (e.g. GENOTYPER and SEQUENCE NAVIGATOR, Perkin-Elmer) may convert the output/light intensity to electrical signal, and the entire process from loading of samples, and computer analysis and electronic data display may be computer controlled. This procedure may be particularly useful for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In accordance with one aspect of the invention, a nucleic acid is provided comprising a *hzf* or *hhl* regulatory sequence such as a promoter sequence.

The invention contemplates nucleic acid molecules comprising all or a portion of a nucleic acid molecule of the invention comprising a regulatory sequence of a *Hzf or Hhl* contained in appropriate vectors. The vectors may contain heterologous nucleic acid sequences. "Heterologous nucleic acid" refers to a nucleic acid not naturally located in the cell, or in a chromosomal site of the cell. Preferably, the heterologous nucleic acid includes a nucleic acid foreign to the cell.

In accordance with another aspect of the invention, the nucleic acid molecules isolated using the methods described herein are mutant *hzf* or *hhl* gene alleles. For example, the mutant alleles may be isolated from individuals either known or proposed to have a genotype which contributes to the symptoms of a hematopoietic disorder. Mutant alleles and mutant allele products may be used in therapeutic and diagnostic methods described herein. For example, a cDNA of a mutant *Hzf or Hhl gene* may be isolated using PCR as described herein, and the DNA sequence of the mutant allele may be compared to the normal allele to ascertain the mutation(s) responsible for the loss or alteration of function of the mutant gene product. A

genomic library can also be constructed using DNA from an individual suspected of or known to carry a mutant allele, or a cDNA library can be constructed using RNA from tissue known, or suspected to express the mutant allele. A nucleic acid encoding a normal *Hzf or Hhl gene* or any suitable fragment thereof, may then be labeled and used as a probe to identify the corresponding mutant allele in such libraries. Clones containing mutant sequences can be purified and subjected to sequence analysis. In addition, an expression library can be constructed using cDNA from RNA isolated from a tissue of an individual known or suspected to express a mutant *hzf* or *hhl* allele. Gene products from putatively mutant tissue may be expressed and screened, for example using antibodies specific for a Hzf or Hhl Polypeptide, or a Hzf or Hhl Related Polypeptide as described herein. Library clones identified using the antibodies can be purified and subjected to sequence analysis.

Antisense molecules and ribozymes are contemplated within the scope of the invention. They may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA sequences encoding Hzf or Hhl Polypeptide. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize antisense RNA constitutively or inducibly can be introduced into cell lines, cells, or tissues. RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Polypeptides of the Invention

The polypeptides of the invention are primarily expressed in hematopoietic lineages. Hzf contains three C_2H_2-type zinc fingers and it is primarily expressed in megakaryocytes and multipotential progenitors. Hhl is expressed in most myeloid lineages with particularly high expression in multipotential progenitors, erythroid, and megakaryocytic cells.

Amino acid sequences of polypeptides of the invention comprise the sequences of SEQ. ID. NO.2., or SEQ. ID. NO. 4. In addition to the amino acid sequences as shown SEQ. ID. NO.2, or SEQ. ID. NO. 4, the polypeptides of the present invention include truncations of the polypeptides of the invention, and analogs, and homologs of the polypeptides and truncations thereof as described herein.

Truncated polypeptides may comprise peptides having an amino acid sequence of at least five consecutive amino acids in SEQ.ID. NO. 2 or 4 where no amino acid sequence of five or more, six or more, seven or more, or eight or more, consecutive amino acids present in the fragment is present in a polypeptide other than a Hzf or Hhl Polypeptide. In an embodiment of the invention the fragment is a stretch of amino acid residues of at least 10 to 100, preferably at least 10 to 75, more preferably 10 to 50, and most

preferably 12 to 20 contiguous amino acids from particular sequences such as the sequences shown in SEQ.ID. NO. 2 or 4. The fragments may be immunogenic and preferably are not immunoreactive with antibodies that are immunoreactive to polypeptides other than Hzf or Hhl.

The truncated polypeptides may have an amino group (-NH2), a hydrophobic group (for example, carbobenzoxyl, dansyl, or T-butyloxycarbonyl), an acetyl group, a 9-fluorenylmethoxy-carbonyl (PMOC) group, or a macromolecule including but not limited to lipid-fatty acid conjugates, polyethylene glycol, or carbohydrates at the amino terminal end. The truncated polypeptides may have a carboxyl group, an amido group, a T-butyloxycarbonyl group, or a macromolecule including but not limited to lipid-fatty acid conjugates, polyethylene glycol, or carbohydrates at the carboxy terminal end.

The polypeptides of the invention may also include analogs, and/or truncations thereof as described herein, which may include, but are not limited to the polypeptides, containing one or more amino acid substitutions, insertions, and/or deletions. Amino acid substitutions may be of a conserved or non-conserved nature. Conserved amino acid substitutions involve replacing one or more amino acids with amino acids of similar charge, size, and/or hydrophobicity characteristics. When only conserved substitutions are made the resulting analog should be functionally equivalent to the native polypeptide. Non-conserved substitutions involve replacing one or more amino acids with one or more amino acids which possess dissimilar charge, size, and/or hydrophobicity characteristics.

One or more amino acid insertions may be introduced into a polypeptide of the invention. Amino acid insertions may consist of single amino acid residues or sequential amino acids ranging from about 2 to 15 amino acids in length.

Deletions may consist of the removal of one or more amino acids, or discrete portions from the polypeptide sequence. The deleted amino acids may or may not be contiguous. The lower limit length of the resulting analog with a deletion mutation is about 10 amino acids, preferably 100 amino acids.

An allelic variant at the polypeptide level differs from another polypeptide by only one, or at most, a few amino acid substitutions. A species variation of a polypeptide of the invention is a variation which is naturally occurring among different species of an organism.

The polypeptides of the invention also include homologs and/or truncations thereof as described herein. Such homologs include polypeptides whose amino acid sequences are comprised of the amino acid sequences of regions from other species that hybridize under selective hybridization conditions (see discussion of selective and in particular stringent hybridization conditions herein) with a probe used to obtain a polypeptide of the invention. These homologs will generally have the same regions which are characteristic of a polypeptide of the invention. It is anticipated that a polypeptide comprising an amino acid sequence which is at least 75% identity or at least 80% similarity, preferably 80 to 90% identity or 90% similarity, more preferably 90 to 95% identity or 95% similarity, and most preferably 95 to 99% identity or 99% similarity with an amino acid sequence of SEQ. ID. NO.2 or SEQ. ID. NO 4 will be a homolog. A percent amino acid sequence similarity or identity is calculated using the methods described herein, preferably the computer programs described herein. For example, a percent amino acid sequence homology or identity is calculated as the percentage of aligned amino acids that match the reference

sequence, where the sequence alignment has been determined using the alignment algorithm of Dayhoff et al; Methods in Enzymology 91: 524-545 (1983).

The invention also contemplates isoforms of the polypeptides of the invention. An isoform contains the same number and kinds of amino acids as the polypeptide of the invention, but the isoform has a different molecular structure. The isoforms contemplated by the present invention are those having the same properties as a polypeptide of the invention as described herein.

The present invention also includes polypeptides of the invention conjugated with a selected polypeptide, or a marker polypeptide (see below) to produce fusion polypeptides. Additionally, immunogenic portions of a polypeptide of the invention are within the scope of the invention.

A polypeptide of the invention may be prepared using recombinant DNA methods. Accordingly, the nucleic acid molecules of the present invention having a sequence which encodes a polypeptide of the invention may be incorporated in a known manner into an appropriate expression vector which ensures good expression of the polypeptide. Possible expression vectors include but are not limited to cosmids, plasmids, or modified viruses (e.g. replication defective retroviruses, adenoviruses and adeno-associated viruses), so long as the vector is compatible with the host cell used.

The invention therefore contemplates a vector of the invention containing a nucleic acid molecule of the invention, and optionally the necessary regulatory sequences for the transcription and translation of the inserted polypeptide-sequence. Suitable regulatory sequences may be derived from a variety of sources, including bacterial, fungal, viral, mammalian, or insect genes (For example, see the regulatory sequences described in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990). Selection of appropriate regulatory sequences is dependent on the host cell chosen as discussed below, and may be readily accomplished by one of ordinary skill in the art. The necessary regulatory sequences may be supplied by a native polypeptide and/or its flanking regions.

The invention further provides a vector comprising a DNA nucleic acid molecule of the invention cloned into the vector in an antisense orientation. That is, the DNA molecule is linked to a regulatory sequence in a manner which allows for expression, by transcription of the DNA molecule, of an RNA molecule which is antisense to a nucleic acid sequence of a nucleic acid molecule of the invention. Regulatory sequences linked to the antisense nucleic acid can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance a viral promoter and/or enhancer, or regulatory sequences can be chosen which direct tissue or cell type specific expression of antisense RNA.

The expression vector of the invention may also contain a marker gene which facilitates the selection of host cells transformed or transfected with a vector of the invention. Examples of marker genes are genes encoding a polypeptide such as G418 and hygromycin which confer resistance to certain drugs, β-galactosidase, chloramphenicol acetyltransferase, firefly luciferase, or an immunoglobulin or portion thereof such as the Fc portion of an immunoglobulin preferably IgG. The markers can be introduced on a separate vector from the nucleic acid of interest.

The vectors may also contain genes which encode a fusion moiety which provides increased

expression of the recombinant polypeptide; increased solubility of the recombinant polypeptide; and aid in the purification of the target recombinant polypeptide by acting as a ligand in affinity purification. For example, a proteolytic cleavage site may be added to the target recombinant polypeptide to allow separation of the recombinant polypeptide from the fusion moiety subsequent to purification of the fusion polypeptide. Typical fusion expression vectors include pGEX (Amrad Corp., Melbourne, Australia), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding polypeptide, or polypeptide A, respectively, to the recombinant polypeptide.

The vectors may be introduced into host cells to produce a transformed or transfected host cell. The terms "transfected " and "transfection" encompass the introduction of nucleic acid (e.g. a vector) into a cell by one of many standard techniques. A cell is "transformed" by a nucleic acid when the transfected nucleic acid effects a phenotypic change. Prokaryotic cells can be transfected or transformed with nucleic acid by, for example, electroporation or calcium-chloride mediated transformation. Nucleic acid can be introduced into mammalian cells via conventional techniques such as calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofectin, electroporation or microinjection. Suitable methods for transforming and transfecting host cells can be found in Sambrook et al. (Molecular Cloning: A Laboratory Manual, 2nd Edition, Cold Spring Harbor Laboratory press (1989)), and other laboratory textbooks.

Suitable host cells include a wide variety of prokaryotic and eukaryotic host cells. For example, the polypeptides of the invention may be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus), yeast cells, or mammalian cells. Other suitable host cells can be found in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (199 1).

A host cell may also be chosen which modulates the expression of an inserted nucleic acid sequence, or modifies (e.g. glycosylation or phosphorylation) and processes (e.g. cleaves) the polypeptide in a desired fashion. Host systems or cell lines may be selected which have specific and characteristic mechanisms for post-translational processing and modification of polypeptides. For example, eukaryotic host cells including CHO, VERO, BHK, HeLA, COS, MDCK, 293, 3T3, and WI38 may be used. For long-term high-yield stable expression of the polypeptide, cell lines and host systems which stably express the gene product may be engineered.

Host cells and in particular cell lines produced using the methods described herein may be particularly useful in screening and evaluating substances or compounds that modulate the activity of a polypeptide of the invention.

The polypeptides of the invention may also be expressed in non-human transgenic animals including but not limited to mice, rats, rabbits, guinea pigs, micro-pigs, goats, sheep, pigs, non-human primates (e.g. baboons, monkeys, and chimpanzees) (see Hammer et al. (Nature 315:680-683, 1985), Palmiter et al. (Science 222:809-814, 1983), Brinster et al. (Proc Natl. Acad. Sci USA 82:44384442, 1985), Palmiter and Brinster (Cell. 41:343-345, 1985) and U.S. Patent No. 4,736,866). Procedures known in the art may be used to introduce a nucleic acid molecule of the invention encoding a polypeptide of the

invention into animals to produce the founder lines of transgenic animals. Such procedures include pronuclear microinjection, retrovirus mediated gene transfer into germ lines, gene targeting in embryonic stem cells, electroporation of embryos, and sperm-mediated gene transfer.

The present invention contemplates a transgenic animal that carries a nucleic acid molecule of the invention in all their cells, and animals which carry the transgene in some but not all their cells. The transgene may be integrated as a single transgene or in concatamers. The transgene may be selectively introduced into and activated in specific cell types (See for example, Lasko et al, 1992 Proc. Natl. Acad. Sci. USA 89: 6236). The transgene may be integrated into the chromosomal site of the endogenous gene by gene targeting. The transgene may be selectively introduced into a particular cell type inactivating the endogenous gene in that cell type (See Gu et al Science 265: 103-106).

The expression of a recombinant polypeptide of the invention in a transgenic animal may be assayed using standard techniques. Initial screening may be conducted by Southern Blot analysis, or PCR methods to analyze whether the transgene has been integrated. The level of mRNA expression in the tissues of transgenic animals may also be assessed using techniques including Northern blot analysis of tissue samples, *in situ* hybridization, and RT-PCR. Tissue may also be evaluated immunocytochemically using antibodies against Hzf or Hhl Polypeptide.

The polypeptides of the invention may also be prepared by chemical synthesis using techniques well known in the chemistry of polypeptides such as solid phase synthesis (Merrifield, 1964, J. Am. Chem. Assoc. 85:2149-2154), or synthesis in homogenous solution (Houbenweyl, 1987, Methods of Organic Chemistry, ed. E. Wansch, Vol. 15 I and II, Thieme, Stuttgart).

N-terminal or C-terminal fusion or chimeric polypeptides comprising a polypeptide of the invention conjugated with other molecules, such as polypeptides m (e.g. markers) may be prepared by fusing, through recombinant techniques, the N-terminal or C-terminal of a polypeptide of the invention, and the sequence of a selected polypeptide or marker polypeptide with a desired biological function. The resultant fusion polypeptides contain a polypeptide of the invention fused to the selected polypeptide or marker polypeptide as described herein. Examples of polypeptides which may be used to prepare fusion polypeptides include immunoglobulins, glutathione-S-transferase (GST), hemagglutinin (HA), and truncated myc.

Antibodies

A polypeptide of the invention can be used to prepare antibodies specific for the polypeptides. Antibodies can be prepared which bind a distinct epitope in an unconserved region of the polypeptide. An unconserved region of the polypeptide is one that does not have substantial sequence homology to other polypeptides. A region from a conserved region such as a well-characterized sequence can also be used to prepare an antibody to a conserved region of a polypeptide of the invention. Antibodies having specificity for a polypeptide of the invention may also be raised from fusion polypeptides created by expressing fusion polypeptides in host cells as described herein.

The invention can employ intact monoclonal or polyclonal antibodies, and immunologically active fragments (e.g. a Fab or $(Fab)_2$ fragment), an antibody heavy chain, and antibody light chain, a genetically

engineered single chain F_v molecule (Ladner et al, U.S. Pat. No. 4,946,778), or a chimeric antibody, for example, an antibody which contains the binding specificity of a murine antibody, but in which the remaining portions are of human origin. Antibodies including monoclonal and polyclonal antibodies, fragments and chimeras, may be prepared using methods known to those skilled in the art.

Applications of the Nucleic Acid Molecules, Polypeptides, and Antibodies of the Invention

The nucleic acid molecules, Hzf or Hhl Polypeptide, or a Hzf or Hhl Related Polypeptide, and antibodies of the invention may be used in the prognostic and diagnostic evaluation of conditions requiring modulation of a nucleic acid or polypeptide of the invention, including disorders of the hematopoietic system, and the identification of subjects with a predisposition to such conditions (See below). Methods for detecting nucleic acid molecules and polypeptides of the invention can be used to monitor such conditions by detecting and localizing the nucleic acids and polypeptides. It would also be apparent to one skilled in the art that the methods described herein may be used to study the developmental expression of the polypeptides of the invention and, accordingly, will provide further insight into the role of the polypeptides. The applications of the present invention also include methods for the identification of substances or compounds that modulate the biological activity of a polypeptide of the invention (See below). The substances, compounds, antibodies etc may be used for the treatment of conditions requiring modulation of polypeptides of the invention including hematopoietic disorders. (See below).

Diagnostic Methods

A variety of methods can be employed for the diagnostic and prognostic evaluation of conditions requiring modulation of a nucleic acid or polypeptide of the invention (e.g. hematopoietic disorders), and the identification of subjects with a predisposition to such conditions. Such methods may, for example, utilize nucleic acid molecules of the invention, and fragments thereof, and antibodies directed against polypeptides of the invention, including peptide fragments. In particular, the nucleic acids and antibodies may be used, for example, for: (1) the detection of the presence of *Hzf or Hhl* mutations, or the detection of either over- or under-expression of *hzf* or *hhl* mRNA relative to a non-disorder state or the qualitative or quantitative detection of alternatively spliced forms of h*zf or hhl* transcripts which may correlate with certain conditions or susceptibility toward such conditions; or (2) the detection of either an over- or an under-abundance of a polypeptide of the invention relative to a non-disorder state or the presence of a modified (e.g., less than full length) polypeptide of the invention which correlates with a disorder state, or a progression toward a disorder state.

The methods described herein may be performed by utilizing pre-packaged diagnostic kits comprising at least one specific nucleic acid or antibody described herein, which may be conveniently used, e.g., in clinical settings, to screen and diagnose patients and to screen and identify those individuals exhibiting a predisposition to developing a disorder.

Nucleic acid-based detection techniques and peptide detection techniques are described below. The samples that may be analyzed using the methods of the invention include those which are known or suspected to express *hzf or hhl* or contain a polypeptide of the invention. The methods may be performed on biological samples including but not limited to cells, lysates of cells which have been incubated in cell

culture, chromosomes isolated from a cell (e.g. a spread of metaphase chromosomes), genomic DNA (in solutions or bound to a solid support such as for Southern analysis), RNA (in solution or bound to a solid support such as for northern analysis), cDNA (in solution or bound to a solid support), an extract from cells or a tissue, and biological fluids such as serum, urine, blood, and CSF. The samples may be derived from a patient or a culture.

The polypeptides of the invention are primarily expressed in hematopoietic lineages, and in particular in megakaryocytes. The polypeptides of the invention have a role in proliferation, differentiation, activation and/or metabolism of cells of the hematopoietic lineages. Therefore, the methods described herein for detecting nucleic acid molecules and polypeptides can be used to monitor proliferation, differentiation, activation and/or metabolism of cells of the hematopoietic lineage (e.g. megakaryocytes) by detecting and localizing polypeptides and nucleic acid molecules of the invention. The methods described herein may be used to study the developmental expression of a polypeptide of the invention and, accordingly, will provide further insight into the role of the polypeptide in the hematopoietic system.

The nucleic acid molecules and polypeptides of the invention are markers for hematopoietic cells and accordingly the antibodies and probes described herein may be used to label these cells.

Methods for Detecting Nucleic Acid Molecules of the Invention

The nucleic acid molecules of the invention allow those skilled in the art to construct nucleotide probes for use in the detection of nucleic acid sequences of the invention in biological materials. Suitable probes include nucleic acid molecules based on nucleic acid sequences encoding at least 5 sequential amino acids from regions of the Hzf or Hhl Polypeptide, or a Hzf or Hhl Related Polypeptide (see SEQ. ID. No. 1 or 3), preferably they comprise 15 to 30 nucleotides. Examples of probes are shown in SEQ. ID. NOs. 5 to 14.

A nucleotide probe may be labeled with a detectable substance such as a radioactive label that provides for an adequate signal and has sufficient half-life such as ^{32}P, ^{3}H, ^{14}C or the like. Other detectable substances that may be used include antigens that are recognized by a specific labeled antibody, fluorescent compounds, enzymes, antibodies specific for a labeled antigen, and luminescent compounds. An appropriate label may be selected having regard to the rate of hybridization and binding of the probe to the nucleotide to be detected and the amount of nucleotide available for hybridization. Labeled probes may be hybridized to nucleic acids on solid supports such as nitrocellulose filters or nylon membranes as generally described in Sambrook et al, 1989, Molecular Cloning, A Laboratory Manual (2nd ed.). The nucleic acid probes may be used to detect *Hzf or Hhl* genes, preferably in human cells. The nucleotide probes may also be useful for example in the diagnosis or prognosis of conditions such as hematopoietic disorders, and in monitoring the progression of these conditions, or monitoring a therapeutic treatment.

The probe may be used in hybridization techniques to detect a *hzf or hhl* gene. The technique generally involves contacting and incubating nucleic acids (e.g. recombinant DNA molecules, cloned genes) obtained from a sample from a patient or other cellular source with a probe of the present invention under conditions favourable for the specific annealing of the probes to complementary sequences in the nucleic acids. After incubation, the non-annealed nucleic acids are removed, and the presence of nucleic

acids that have hybridized to the probe if any are detected.

The detection of nucleic acid molecules of the invention may involve the amplification of specific gene sequences using an amplification method such as PCR, followed by the analysis of the amplified molecules using techniques known to those skilled in the art. Suitable primers can be routinely designed by one of skill in the art.

Genomic DNA may be used in hybridization or amplification assays of biological samples to detect abnormalities involving *hzf or hhl* structure, including point mutations, insertions, deletions, and chromosomal rearrangements. For example, direct sequencing, single stranded conformational polymorphism analyses, heteroduplex analysis, denaturing gradient gel electrophoresis, chemical mismatch cleavage, and oligonucleotide hybridization may be utilized.

Genotyping techniques known to one skilled in the art can be used to type polymorphisms that are in close proximity to the mutations in a *hzf or hhl* gene. The polymorphisms may be used to identify individuals in families that are likely to carry mutations. If a polymorphism exhibits linkage disequalibrium with mutations in the *hzf or hhl* gene, it can also be used to screen for individuals in the general population likely to carry mutations. Polymorphisms which may be used include restriction fragment length polymorphisms (RFLPs), single-base polymorphisms, and simple sequence repeat polymorphisms (SSLPs).

A probe of the invention may be used to directly identify RFLPs. A probe or primer of the invention can additionally be used to isolate genomic clones such as YACs, BACs, PACs, cosmids, phage or plasmids. The DNA in the clones can be screened for SSLPs using hybridization or sequencing procedures.

Hybridization and amplification techniques described herein may be used to assay qualitative and quantitative aspects of *hzf or hhl* expression. For example, RNA may be isolated from a cell type or tissue known to express *hzf or hhl* and tested utilizing the hybridization (e.g. standard Northern analyses) or PCR techniques referred to herein. The techniques may be used to detect differences in transcript size that may be due to normal or abnormal alternative splicing. The techniques may be used to detect quantitative differences between levels of full length and/or alternatively splice transcripts detected in normal individuals relative to those individuals exhibiting symptoms of a disease.

The primers and probes may be used in the above-described methods *in situ* i.e directly on tissue sections (fixed and/or frozen) of patient tissue obtained from biopsies or resections.

Oligonucleotides or longer fragments derived from any of the nucleic acid molecules of the invention may be used as targets in a microarray. The microarray can be used to simultaneously monitor the expression levels of large numbers of genes and to identify genetic variants, mutations, and polymorphisms. The information from the microarray may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, and to develop and monitor the activities of therapeutic agents.

The preparation, use, and analysis of microarrays are well known to a person skilled in the art. (See, for example, Brennan, T. M. et al. (1995) U.S. Pat. No. 5,474,796; Schena, et al. (1996) Proc. Natl. Acad. Sci. 93:10614-10619; Baldeschweiler et al. (1995), PCT Application WO95/251116; Shalon, D. et

al. (I 995) PCT application WO95/35505; Heller, R. A. et al. (1997) Proc. Natl. Acad. Sci. 94:2150-2155; and Heller, M. J. et al. (1997) U.S. Pat. No. 5.605,662.)

Methods for Detecting Polypeptides

Antibodies specifically reactive with a Hzf or Hhl Polypeptide, a Hzf or Hhl Related Polypeptide, or derivatives, such as enzyme conjugates or labeled derivatives, may be used to detect Hzf or Hhl Polypeptides or Hzf or Hhl Related Polypeptides in various biological materials. They may be used as diagnostic or prognostic reagents and they may be used to detect abnormalities in the level of Hzf or Hhl Polypeptides or Hzf or Hhl Related Polypeptides, expression, or abnormalities in the structure, and/or temporal, tissue, cellular, or subcellular location of the polypeptides. Antibodies may also be used to screen potentially therapeutic compounds *in vitro* to determine their effects on a condition such as a hematopoietic disorder. *In vitro* immunoassays may also be used to assess or monitor the efficacy of particular therapies. The antibodies of the invention may also be used *in vitro* to determine the level of Hzf or Hhl Polypeptide or Hzf or Hhl Related Polypeptide expression in cells genetically engineered to produce a Hzf or Hhl Polypeptide, or Hzf or Hhl Related Polypeptide.

The antibodies may be used in any known immunoassays that rely on the binding interaction between an antigenic determinant of a polypeptide of the invention, and the antibodies. Examples of such assays are radioimmunoassays, enzyme immunoassays (e.g. ELISA), immunofluorescence, immunoprecipitation, latex agglutination, hemagglutination, and histochemical tests. The antibodies may be used to detect and quantify polypeptides of the invention in a sample in order to determine their role in particular cellular events or pathological states, and to diagnose and treat such pathological states.

In particular, the antibodies of the invention may be used in immuno-histochemical analyses, for example, at the cellular and sub-subcellular level, to detect a polypeptide of the invention, to localise it to particular cells and tissues, and to specific subcellular locations, and to quantitate the level of expression.

Cytochemical techniques known in the art for localizing antigens using light and electron microscopy may be used to detect a polypeptide of the invention. Generally, an antibody of the invention may be labeled with a detectable substance and a polypeptide may be localised in tissues and cells based upon the presence of the detectable substance. Various methods of labeling polypeptides are known in the art and may be used. Examples of detectable substances include, but are not limited to, the following: radioisotopes (e.g., ^{3}H, ^{14}C, ^{35}S, ^{125}I, ^{131}I), fluorescent labels (e.g., FITC, rhodamine, lanthanide phosphors), luminescent labels such as luminol; enzymatic labels (e.g., horseradish peroxidase, β-galactosidase, luciferase, alkaline phosphatase, acetylcholinesterase), biotinyl groups (which can be detected by marked avidin e.g., streptavidin containing a fluorescent marker or enzymatic activity that can be detected by optical or calorimetric methods), predetermined polypeptide epitopes recognized by a secondary reporter (e.g., leucine zipper pair sequences, binding sites for secondary antibodies, metal binding domains, epitope tags). In some embodiments, labels are attached via spacer arms of various lengths to reduce potential steric hindrance. Antibodies may also be coupled to electron dense substances, such as ferritin or colloidal gold, which are readily visualised by electron microscopy.

The antibody or sample may be immobilized on a carrier or solid support which is capable of

immobilizing cells, antibodies, etc. For example, the carrier or support may be nitrocellulose, or glass, polyacrylamides, gabbros, and magnetite. The support material may have any possible configuration including spherical (e.g. bead), cylindrical (e.g. inside surface of a test tube or well, or the external surface of a rod), or flat (e.g. sheet, test strip). Indirect methods may also be employed in which the primary antigen-antibody reaction is amplified by the introduction of a second antibody, having specificity for the antibody reactive against a polypeptide of the invention. By way of example, if the antibody having specificity against a polypeptide of the invention is a rabbit IgG antibody, the second antibody may be goat anti-rabbit gamma-globulin labeled with a detectable substance as described herein.

Where a radioactive label is used as a detectable substance, a polypeptide of the invention may be localized by radioautography. The results of radioautography may be quantitated by determining the density of particles in the radioautographs by various optical methods, or by counting the grains.

Methods for Identifying or Evaluating Substances/Compounds

The methods described herein are designed to identify substances or compounds that modulate the biological activity of a Hzf or HHl Polypeptide or Hzf or HHl Related Polypeptide. "Modulate" refers to a change or an alteration in the biological activity of a polypeptide of the invention. Modulation may be an increase or a decrease in activity, a change in characteristics, or any other change in the biological, functional, or immunological properties of the polypeptide.

Substances and compounds identified using the methods of the invention include but are not limited to peptides such as soluble peptides including Ig-tailed fusion peptides, members of random peptide libraries and combinatorial chemistry-derived molecular libraries made of D- and/or L-configuration amino acids, phosphopeptides (including members of random or partially degenerate, directed phosphopeptide libraries), antibodies [e.g. polyclonal, monoclonal, humanized, anti-idiotypic, chimeric, single chain antibodies, fragments, (e.g. Fab, $F(ab)_2$, and Fab expression library fragments, and epitope-binding fragments thereof)], and small organic or inorganic molecules. A substance or compound may be an endogenous physiological compound or it may be a natural or synthetic compound.

Substances which modulate a Hzf or Hhl Polypeptide or Hzf or Hhl Related Polypeptide can be identified based on their ability to associate with a Hzf or Hhl Polypeptide or Hzf or Hhl Related Polypeptide. Therefore, the invention also provides methods for identifying substances which associate with a Hzf or Hhl Polypeptide or Hzf or Hhl Related Polypeptide. Substances identified using the methods of the invention may be isolated, cloned and sequenced using conventional techniques. A substance that associates with a polypeptide of the invention may be an agonist or antagonist of the biological or immunological activity of a polypeptide of the invention.

The term "agonist", refers to a molecule that increases the amount of, or prolongs the duration of, the activity of the polypeptide. The term "antagonist" refers to a molecule which decreases the biological or immunological activity of the polypeptide. Agonists and antagonists may include proteins, nucleic acids, carbohydrates, or any other molecules that associate with a polypeptide of the invention.

Substances which can associate with a polypeptide of the invention may be identified by reacting the polypeptide with a test substance which potentially associates with the polypeptide, under conditions

which permit the association, and removing and/or detecting the associated polypeptides and substance. Substance-polypeptide complexes, free substance, non-complexed polypeptide, or activated polypeptide may be assayed. Conditions which permit the formation of complexes may be selected having regard to factors such as the nature and amounts of the substance and the polypeptide.

Substance-polypeptide complexes, free substances or non-complexed polypeptides may be isolated by conventional isolation techniques, for example, salting out, chromatography, electrophoresis, gel filtration, fractionation, absorption, polyacrylamide gel electrophoresis, agglutination, or combinations thereof. To facilitate the assay of the components, antibody against the polypeptide or the substance, or labelled polypeptide, or a labelled substance may be utilized. The antibodies, polypeptides, or substances may be labelled with a detectable substance as described above.

A polypeptide, or the substance used in the method of the invention may be insolubilized. For example, the polypeptide, or substance may be bound to a suitable carrier such as agarose, cellulose, dextran, Sephadex, Sepharose, carboxymethyl cellulose polystyrene, filter paper, ion-exchange resin, plastic film, plastic tube, glass beads, polyamine-methyl vinyl-ether-maleic acid copolymer, amino acid copolymer, ethylene-maleic acid copolymer, nylon, silk, etc. The carrier may be in the shape of, for example, a tube, test plate, beads, disc, sphere etc. The insolubilized polypeptide or substance may be prepared by reacting the material with a suitable insoluble carrier using known chemical or physical methods, for example, cyanogen bromide coupling.

The invention also contemplates a method for evaluating a compound for its ability to modulate the biological activity of a polypeptide of the invention, by assaying for an agonist or antagonist (i.e. enhancer or inhibitor) of the association of the polypeptide with a substance that associates with the polypeptide. The basic method for evaluating if a compound is an agonist or antagonist of the association of a polypeptide of the invention and a substance that associates with the polypeptide, is to prepare a reaction mixture containing the polypeptide and the substance under conditions which permit the formation of substance- polypeptide complexes, in the presence of a test compound. The test compound may be initially added to the mixture, or may be added subsequent to the addition of the polypeptide and substance. Control reaction mixtures without the test compound or with a placebo are also prepared. The formation of complexes is detected and the formation of complexes in the control reaction but not in the reaction mixture indicates that the test compound interferes with the interaction of the polypeptide and substance. The reactions may be carried out in the liquid phase or the polypeptide, substance, or test compound may be immobilized as described herein.

It will be understood that the agonists and antagonists i.e. inhibitors and enhancers that can be assayed using the methods of the invention may act on one or more of the binding sites on the polypeptide or substance including agonist binding sites, competitive antagonist binding sites, non-competitive antagonist binding sites, or allosteric sites.

The invention also makes it possible to screen for antagonists that inhibit the effects of an agonist of the interaction of the polypeptide with a substance which is capable of binding to the polypeptide. Thus, the invention may be used to assay for a compound that competes for the same binding site of the

polypeptide.

The reagents suitable for applying the methods of the invention to evaluate compounds that modulate a polypeptide of the invention may be packaged into convenient kits providing the necessary materials packaged into suitable containers. The kits may also include suitable supports useful in performing the methods of the invention.

Compositions and Treatments

The polypeptides, nucleic acid molecules, substances or compounds identified by the methods described herein, antibodies, and antisense nucleic acid molecules of the invention may be used for modulating the biological activity of a polypeptide or nucleic acid molecule of the invention. The polypeptides etc. may have particular application in the treatment of conditions requiring modulation of cells of the hematopoietic lineage i.e. hematopoietic disorders. Treatable disorders include anemia, thrombocytopenia, leukopenia, myelofibrosis, hypoplasia, disseminated intravascular coagulation, immune, thromocytopenic purpura, myelodysplasia, erythrocytopenia, erthordegenerative disorders, erythroblastopenia, leukoerythroblastosis, erythroclasis, thalassemia, and granulocytopenia. The polypeptides etc. may be used to modulate production of blood cells in situations where a patient has a disease such as AIDS, which has caused alteration in normal blood cell levels, or in clinical settings in which enhancement of hematopoietic populations is desired, such as in conjunction with bone marrow transplant, or in the treatment of aplasia or myelosuppression caused by radiation, chemical treatment, or chemotherapy. The polypeptides, nucleic acids, etc. may also be used to treat patients in which the desired result is the inhibition of a hematopoietic pathway such as for the treatment of myeloproliferative or other proliferative disorders of blood forming organs such as thromocythemias, polycythemias, and leukemias.

The polypeptides, nucleic acids, etc. may have specific application in the regulation of megakaryocytopoiesis. Abnormal enhancement of megakaryocytopoiesis leads to thrombocytosis which is implicated in a number of debilitating blood cell disorders, including leukemias and carcinomas. Therefore, the polypeptides, nucleic acids, etc. may be used in the prevention and treatment of polycythemiz vera, megakaryocytic leukemia, ovarian adenocarcinoma, and lupus erythematosus, and in the prevention and treatment of many common diseases where functioning of platelets requires modulation (e.g. inhibition) such as thrombosis, atherosclerosis, hemorrhage, embolism, myocardial infarction, post-myocardial infarction syndrome, and post-cardiac surgery.

Accordingly, the polypeptides, substances, antibodies, and compounds may be formulated into pharmaceutical compositions for adminstration to subjects in a biologically compatible form suitable for administration *in vivo*. By "biologically compatible form suitable for administration *in vivo*" is meant a form of the substance to be administered in which any toxic effects are outweighed by the therapeutic effects.. The substances may be administered to living organisms including humans, and animals. Administration of a therapeutically active amount of the pharmaceutical compositions of the present invention is defined as an amount effective, at dosages and for periods of time necessary to achieve the desired result. For example, a therapeutically active amount of a substance may vary according to factors such as the disease state, age, sex, and weight of the individual, and the ability of antibody to elicit a desired

response in the individual. Dosage regima may be adjusted to provide the optimum therapeutic response. For example, several divided doses may be administered daily or the dose may be proportionally reduced as indicated by the exigencies of the therapeutic situation.

The active substance may be administered in a convenient manner such as by injection (subcutaneous, intravenous, etc.), oral administration, inhalation, transdermal application, or rectal administration. Depending on the route of administration, the active substance may be coated in a material to protect the compound from the action of enzymes, acids and other natural conditions which may inactivate the compound.

The compositions described herein can be prepared by per se known methods for the preparation of pharmaceutically acceptable compositions which can be administered to subjects, such that an effective quantity of the active substance is combined in a mixture with a pharmaceutically acceptable vehicle. Suitable vehicles are described, for example, in Remington's Pharmaceutical Sciences (Remington's Pharmaceutical Sciences, Mack Publishing Company, Easton, Pa., USA 1985). On this basis, the compositions include, albeit not exclusively, solutions of the substances or compounds in association with one or more pharmaceutically acceptable vehicles or diluents, and contained in buffered solutions with a suitable pH and iso-osmotic with the physiological fluids.

After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of a polypeptide, nucleic acids, etc. of the invention, such labeling would include amount, frequency, and method of administration.

The polypeptides, nucleic acids, etc. and compositions of the invention may be used alone, or in combination with another pharmaceutically active agent such as, for example, cytokines, neurotrophins, interleukins, etc. For example, the polypeptides etc. and compositions may be used in conjunction with any of a number of the above-referenced factors which are known to induce stem cell or other hematopoietic precursor proliferation, or factors acting on later cells in the hematopoietic pathway including but not limited to hemopoietic maturation factor, thrombopoietin, stem cell factor, erythropoietin, G-CSF, GM-CSF etc.

The invention also contemplates an antibody that specifically binds the therapeutically active ingredient used in a treatment or composition of the invention. The antibody may be used to measure the amount of the therapeutic molecule in a sample taken from a patient for purposes of monitoring the course of therapy.

The nucleic acid molecules encoding a polypeptide of the invention or any fragment thereof, or antisense sequences may be used for therapeutic purposes. Antisense to a nucleic acid molecule encoding a polypeptide of the invention may be used in situation to block the synthesis of the polypeptide. In particular, cells may be transformed with sequences complementary to nucleic acid molecules encoding a Hzf or Hhl Polypeptide or Hzf or Hhl Related Polypeptide. Thus, antisense sequences may be used to modulate activity or to achieve regulation of gene function. This technology is well known in the art, and sense or antisense oligomers or larger fragments, can be designed from various locations along the coding or regulatory regions of sequences encoding a polypeptide of the invention.

Expression vectors may be derived from retroviruses, adenoviruses, herpes or vaccinia viruses or from various bacterial plasmids for delivery of nucleic acid sequences to the target organ, tissue, or cells. Vectors that express antisense nucleic acid sequences of Hzf or Hhl Polypeptides can be constructed using techniques well known to those skilled in the art.

Genes encoding a Hzf or Hhl Polypeptide can be turned off by transforming a cell or tissue with expression vectors that express high levels of a nucleic acid molecule or fragment thereof which encodes a polypeptide of the invention. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even if they do not integrate into the DNA, the vectors may continue to transcribe RNA molecules until all copies are disabled by endogenous nucleases. Transient expression may last for extended periods of time (e.g a month or more) with a non-replicating vector, or if appropriate replication elements are part of the vector system.

Modification of gene expression may be achieved by designing antisense molecules, DNA, RNA, or PNA, to the control regions of a *hzf* or *hhl* gene i.e. the promoters, enhancers, and introns. Preferably the antisense molecules are oligonucleotides derived from the transcription initiation site (e.g. between positions –10 and +10 from the start site). Inhibition can also be achieved by using triple-helix base-pairing techniques. Triple helix pairing causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules (see Gee J.E. et al (1994) In: Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, N.Y.). An antisense molecule may also be designed to block translation of mRNA by inhibiting binding of the transcript to the ribosomes.

Ribozymes, enzymatic RNA molecules, may be used to catalyze the specific cleavage of RNA. Ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, hammerhead motif ribozyme molecules may be engineered that can specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding a polypeptide of the invention.

Specific ribosome cleavage sites within any RNA target may be initially identified by scanning the target molecule for ribozyme cleavage sites which include the following sequences: GUA, GUU, and GUC. Short RNA sequences of between 15 and 20 ribonucleotides corresponding to the region of the cleavage site of the target gene may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The therapeutic index is the dose ratio of therapeutic to toxic effects and it can be expressed as the ED_{50}/LD_{50} ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred.

The invention also provides methods for studying the function of a polypeptide of the invention. Cells, tissues, and non-human animals lacking in expression or partially lacking in expression of a nucleic

acid molecule or gene of the invention may be developed using recombinant expression vectors of the invention having specific deletion or insertion mutations in the gene. A recombinant expression vector may be used to inactivate or alter the endogenous gene by homologous recombination, and thereby create a deficient cell, tissue, or animal.

Null alleles may be generated in cells, such as embryonic stem cells by deletion mutation. A recombinant gene may also be engineered to contain an insertion mutation that inactivates the gene. Such a construct may then be introduced into a cell, such as an embryonic stem cell, by a technique such as transfection, electroporation, injection etc. Cells lacking an intact gene may then be identified, for example by Southern blotting, Northern Blotting, or by assaying for expression of the encoded polypeptide using the methods described herein. Such cells may then be fused to embryonic stem cells to generate transgenic non-human animals deficient in a polypeptide of the invention. Germline transmission of the mutation may be achieved, for example, by aggregating the embryonic stem cells with early stage embryos, such as 8 cell embryos, *in vitro*; transferring the resulting blastocysts into recipient females and; generating germline transmission of the resulting aggregation chimeras. Such a mutant animal may be used to define specific cell populations, developmental patterns and *in vivo* processes, normally dependent on gene expression.

The invention thus provides a transgenic non-human mammal all of whose germ cells and somatic cells contain a recombinant expression vector that inactivates or alters a gene encoding a Hzf or Hhl Polypeptide. In an embodiment the invention provides a transgenic non-human mammal all of whose germ cells and somatic cells contain a recombinant expression vector that inactivates or alters a gene encoding a Hzf Polypeptide resulting in a Hzf associated pathology. Further the invention provides a transgenic non-human mammal which doe not express a Hzf or Hhl polypeptide of the invention. In an embodiment, the invention provides a transgenic non-human mammal which doe not express a Hzf polypeptide of the invention resulting in a Hzf associated pathology.

An Hzf associated pathology refers to a phenotype observed for a Hzf homozygous mutant. The phenotype is similar to the phenotype observed after sustained brain damage resulting from stroke in humans. The phenotype of an Hzf homozygous mutant mouse includes the following characteristics: hemorrhage frequently observed in the brain, smaller size, animals are able to lift their heads only to the right or left; and their bodies tremor when placed on unstable surfaces. The membranes of megakaryocytes from the Hzf homozygous mutant animals are also aberrant.

A transgenic non-human animal includes but is not limited to mouse, rat, rabbit, sheep, hamster, dog, cat, goat, and monkey, preferably mouse.

The invention also provides a transgenic non-human animal assay system which provides a model system for testing for an agent that reduces or inhibits a pathology associated with an Hzf or Hhl Polypeptide, preferably a Hzf associated pathology, comprising:

(a) administering the agent to a transgenic non-human animal of the invention; and

(b) determining whether said agent reduces or inhibits the pathology (e.g. Hzf associated pathology) in the transgenic mouse relative to a transgenic mouse of step (a) which has not been administered the agent.

The agent may be useful in the treatment of hematopoietic disorders or disorders requiring modulation of megakaryocytopoiesis, or it may be used to modulate production of blood cells as discussed herein. The agents may be particularly useful in the treatment and prophylaxis of thrombosis, hemorrhage, embolism, stroke, atherosclerosis, myocardial infarction, post-myocardial infarction syndrome, and post-cardiac surgery. The agents may be incorporated in a pharmaceutical composition as described herein.

A polypeptide of the invention may be used to support the survival, growth, migration, and/or differentiation of cells expressing the polypeptide. Thus, a polypeptide of the invention may be used as a supplement to support, for example hematopoietic cells in culture.

The following non-limiting examples are illustrative of the present invention:

Examples

Example 1

MATERIAL AND METHODS

Cells. R1 embryonic stem cells from the 129/Sv strain [34] were maintained on a layer of mitomycin C-treated embryonic fibroblasts in Dulbecco's modified Eagle's culture medium, supplemented with leukemia inhibitory factor, 15% fetal calf serum, L-glutamine, and β-mercaptoethanol as previously described [35]. The OP9 stromal cell line were cultured in α-MEM containing 20% fetal calf serum [33].

Generation of Trapped ES Cell Lines. The plasmid PT1-ATG (PT1 henceforth) contains the *En-2* splice acceptor site positioned immediately upstream of the *lacZ* reporter gene with an ATG translational start site [36]. The bacterial *neomycin-resistance* (*neo*) gene is driven by the *phosphoglycerate kinase-1* (*PGK-1*) promoter. After electroporation of the ES cells with the PT1 vector and selection in G418, drug-resistant clones were plated into 96-well plates. Upon confluency, the cells were expanded into three 96-well plates. One plate was frozen; cells in the second plate were stained with X-galactosidase (X-gal); and cells in the third plate were used for hematopoietic differentiation on OP9 cells.

Analysis of β-galactosidase Activity. β-galactosidase (β-gal) activity was analyzed by staining with 5-bromo 4-chloro 3-indolyl β-galactoside (X-gal; Boehringer Mannheim, Germany). Briefly, cells were washed once with phosphate buffered saline (PBS) and fixed in 0.2% glutaraldehyde, 2mM $MgCl_2$ and 5mM EGTA in 100mM Na_2HPO_4 for 10 min. at room temperature. Cell cultures were then washed three times for 5 min. each in 100mM Na_2HPO_4, 0.02% NP-40 and 2 mM $MgCl_2$ and stained overnight in X-gal staining solution (1 mg/ml X-gal, 5 mM $K_3Fe(CN)_6$, 5 mM $K_4Fe(CN)_6$, 2mM $MgCl_2$, 0.02% NP-40 in 100mM Na_2HPO_4) directly in 96- or 6-well plates at 37°C. Following X-gal staining, samples were washed overnight at 4°C. Detection of β-gal activity in embryos was performed as described above except the fixative included 1.5% formaldehyde and fixation was performed for between 30-60 min. depending on the size of the embryo and washed 3 times each for 30 min.

Screening Clones for Hematopoietic Expression. Differentiation of ES cells on OP9 stromal cells was performed essentially as previously described [33]. The drug-resistant clones within each row of a 96-well plate which did not demonstrate X-gal staining in undifferentiated cultures (white clones) were pooled. The pooled clones were seeded onto confluent OP9 cell layers in 6-well plates at a density of 10^4 cells/well in

OP9 media. Cultures were trypsinized at day 5, and 10^5 cells/well were transferred onto fresh OP9 stroma. Mesoderm-like colonies and hematopoietic cell clusters were observed on day 5 and day 10, respectively. *LacZ* expression was examined at day 5 and day 10. The pooled clones that demonstrated β-gal activity during the first round of screening were thawed from frozen stocks, and each clone was tested individually for expression of *lacZ* during hematopoietic differentiation. Clones which demonstrated X-gal staining as undifferentiated cells were analysed individually on OP9 cells.

Molecular cloning of the trapped genes by RACE, inverse PCR and cDNA library screening. RNA was prepared from either undifferentiated or differentiated ES cells using Trizol (Gibco/BRL) according to manufacturer's instructions. 5' RACE was performed using the 5' RACE kit (Gibco/BRL), according to manufacturer's instructions with modifications previously described[17]. 5' RACE products were subcloned into the CloneAmp plasmid (Gibco/BRL) and both strands sequenced using the AutoRead Sequencing kit (Pharmacia) and an A.L.F. DNA Sequencer (Pharmacia). Sequences were analyzed by comparison to the non-redundant GenBank and EST databases of the NCBI using the BLASTN program. 3' RACE was performed using a 3' RACE kit (Life Technologies) and two nested primers complimentary to each 5' RACE fragment (see Figure 5). Wild type total RNA from day 9 embryos or mouse brain was used as template. Specific primers were as follows: 12G10-5, 5'-GCAGAGCTCTGGGCGGCGGGT-3'; 12G10-6, 5'-CTCGGGCGTCTGACAGAGTTG-3'; 4H5-1, 5'-CCCGGGACAGAACGACGC-3'; 4H5-2, 5'-CGGAACCCCGGGAGCCAGG-3' (SEQ ID NOs. 5-8, respectively). To obtain a long amplified product, the Expand™ long template PCR system (Boehringer Mannheim) was used. The RACE-PCR products were cloned into the pCR2.1 vector (Invitrogen). The cDNA inserts were sequenced on both strands as previously described. Inverse PCR was performed essential as described by von Melchner *et al.*, (1990)[37]. Genomic DNA from the Hhl ES cell line was digested with PstI and self-ligated at a concentration of 1μg/ml to obtain circular molecules. 20ng of self-ligated circular DNA was used for the PCR using the Expand™ long template PCR system (Boehringer Mannheim), with primer sets of GTlacZ-3 [5'-ATTACGCCAGCTGGCGAAAGG-3'] (SEQ.ID.NO.9) and GTlacZ-4 [5'-GATTGGTGGCGACGACTCCTG-3'] (SEQ.ID.NO.10). Following digestion with PstI and KpnI to isolate flanking sequences just upstream of vector integration sites, the DNA was separated electophoretically on a 0.7% agarose gel and hybridized with a pGEM7 probe.

A 12.5 d.p.c. mouse embryonic cDNA library in the λEXlox vector (mouse strain NIH Swiss; Novagen) was screened by standard procedures [38] using *Hhl* RACE fragments as probes. Phage plaques ($1.0x10^6$) were transferred to nylon filters (Amersham) and hybridized for 2hr at 68°C with the radiolabeled probe using Quikhyb solution (Stratagene). Filters were washed at 60°C: twice for 15 min in 2x SSC and for 1 hour in 0.1x SSC containing 0.1% SDS. Positive clones were plaque purified and the inserts recovered and sequenced as described above. GenBank search and sequence analysis were performed using Fasta, Translate, Bestfit, and Pileup softwares from the Wisconsin Package (Genetics Computer Group, Inc., Madison, WI). The ORFs of Hhl and Hzf were submitted to the National Center for Biotechnology Information XREF internet service (http://www.ncbi.nlm.nih.gov/XREFdb/) where the

sequences were compared to the most recent updates of the EST database. In addition, the sequences were compared to all public databases using the BLAST algorithm.

Generation of Chimeras. The ES cell clones, Hhl and Hzf were aggregated with CD1 diploid embryos as previously described [34] and transferred to foster mothers to generate several strong male chimeras. Chimeric males were bred to CD1 females and tail DNA of F_1 and F_2 offspring was analyzed by southern blotting and hybridization to *En-2/LacZ* and RACE fragment probes. For the Hhl clone, genotyping was performed using densitometry, comparing the intensity of the *LacZ* band to the internal standard, *En-2*. The reliability of genotyping by quantitative Southern blot analysis was confirmed by test breeding [29]. For the Hzf line, EcoRI-digested DNA was hybridized with the RACE probe which detected a polymorphism between the wild-type and mutant alleles. All the analyses described were performed in the CD-1/129Sv hybrid background.

RNA Analysis. Northern blot analysis was carried out according to standard procedures [38]. RT-PCR was performed using an RNA PCR kit (Perkin-Elmer Cetus) using total RNA from the brains of F_2 mice as template. Combinations of specific primers were as follows: 12G10-6; 12G10-14, 5'-TATCTTCAGCTGTGGCTTCCC-3' (SEQ.ID.NO.11); GT-*lacZ*-2A 5'-CCGTCGACTC TGGCGCCGCTGCTCTGTCAG-3'(SEQ.ID.NO.12); for amplification of Hhl products and 4H5-2, 5'-CGGAACCCCGGGAGCCAGG-3'(SEQ.ID.NO.13); 4H5-6, 5'-TCTGGGGATCCTGGAGCTGGAC-3'(SEQ.ID.NO.14); GT-*lacZ*-2A for Hzf products.

Immunofluorescence. Cells were stained with the fluorescent β-gal substrate, fluorescein di-β-D-galactopyranoside (FDG) as previously described [39]. Cells (10^6) were washed twice with phosphate buffered saline (PBS) and resuspended in 40μl PBS containing 5% FCS. After 10 min of preincubation at 37°C, cells were incubated with 40μl prewarmed FDG (2mM in water) for 75 seconds at 37°C. FDG loading was terminated by adding 720μl ice cold PBS containing 5% FCS. Cells were cytospun onto glass slides and examined under a fluorescent microscope, prior to morphological examination by Wright-Giemsa staining. Cell surface expression of CD61 was performed by indirect immunofluorescence. Bone marrow cells were washed in PBS and incubated with anti-CD61 (PharMingen, ON, CA) for 30 minutes, washed twice, then incubated for 30 minutes with fluorescein isothiocyanate (FITC) conjugated mouse anti-hamster IgG monoclonal antibody (PharMingen, ON, CA). Cells were washed twice and analysis was performed using a FACS cell sorter (Becton and Dickinson, Mountain View, CA)

CFU-C Analysis. Bone marrow cells were cultured in methylcellulose containing the growth factors IL-3, IL-6, Epo and SCF (METHOCULT GF M3434; Stemcell Technologies). After 7-16 days, individual colonies were picked and stained with X-gal in a 96-well plate. Colony types were determined morphologically using Wright-Giemsa staining. For analysis of CFU-Mk, bone marrow cells were plated in 0.3% MegaCult serum-free base agarose (StemCell Technologies HCC-4701K) in the presence of recombinant murine IL-3 (20ng/ml), TPO (40ng/ml) and SLF (50ng/ml) [40] and cultured for 18-21 days. The complete culture was stained with X-gal and CFU-Mk were examined microscopically.

RESULTS

Isolation of hematopoietic gene trap clones. The PT1 gene trap vector, which contains a splice acceptor site immediately upstream of a promoterless *lacZ* reporter gene and the *neo* gene driven by the *PGK-1* promoter, was introduced into ES cells (clone R1) by electroporation. After selection of neo^R colonies, clones were replica plated for freezing, analysis of *lacZ* expression by undifferentiated ES cells, and differentiation on OP9 stromal cells. Of 1,350 neo^R ES clones, 95% did not express *lacZ* (white clones) and 5% were *lacZ* positive (blue clones) in the undifferentiated state. Both white and blue clones were allowed to differentiate on OP9 cells and *lacZ* expression was monitored throughout hematopoietic differentiation. Two of the originally white clones (0.2% of the total number of clones) now expressed *lacZ* upon hematopoietic differentiation (Table 1). One of these clones, 7H7, expressed *lacZ* at day 5 in mesoderm-like colonies which was subsequently downregulated during differentiation and not detected in hematopoietic cells by day 10 of culture. The other clone, Hhl, expressed *lacZ* at both the mesodermal and hematopoietic stages of differentiation (Figure 1). Approximately one-third of the hematopoietic clusters derived from Hhl ES cells expressed *lacZ* 12 days after intitiation of co-culture.

Of the 70 clones that expressed *lacZ* in the undifferentiated state (Table 1), 97% (68/70) downregulated *lacZ* expression upon differentiation while two of the originally blue clones, Hzf and 4A11, expressed *lacZ* throughout the differentiation into hematopoietic cells (Table 1). Clone Hzf expressed *lacZ* in only a few cells at the undifferentiated (day 2) and mesodermal stages (day 5). Interestingly, *lacZ* expression in Hzf was restricted to large cells located at the periphery of day 12 hematopoietic clusters (Figure 1). Clone 4A11 was subsequently shown to be ubiquitously expressed *in vivo* and was not characterized further. The two hematopoietic regulated gene trap clones Hzf and Hhl are described below.

***In vivo* Expression Analysis.** Previously it has been shown that the *in vitro* patterns of *lacZ* expression in trapped ES clones accurately predicts the expression patterns observed *in vivo* [18]. To determine whether the *in vitro* expression patterns of the trapped genes in the ES clones Hhl and Hzf were also recapitulated *in vivo,* chimeric mice were generated between these ES clones and diploid embryos and the trapped alleles were transmitted through the germ line to produce F_1 heterozygotes. In Hhl heterozygotes (*Hhl/+*), *lacZ* was expressed in the yolk sac and heart primordium at 8.5 days postcoitum (d.p.c) and in the heart, dorsal root ganglia and fetal liver after 12.5 d.p.c (Figure 2A & B). Virtually all cells in 14.5 d.p.c fetal liver were positive for X-gal staining (Figure 2C), and *lacZ* expression in the heart was localized to the endocardium and myocardium (Figure 2D). In Hzf heterozygotes (*Hzf/+*), *lacZ* was expressed in the somites, basal ganglia, apical ectodermal region of the limb buds and liver primordium in 9.5 d.p.c. embryos (Figure 2E). Expression was also observed in the skin (Figure 2H), trigeminal ganglia, thymus, salivary gland and spinal cord, with punctate staining in the fetal liver around 14.5 d.p.c (Figure 2F). The punctate staining in the liver was due to the restricted staining of large polynuclear cells resembling megakaryocytes (Figure 2G).

To characterize further the expression pattern of the trapped genes within the hematopoietic compartment, bone marrow cells were isolated from adult mice and analyzed for *lacZ* expression.

Unfractionated cells obtained from *Hzfl*+ mice expressed high levels of *lacZ* in about 2% of total bone marrow cells while up to 80% (ranging from very high to low levels of expression) of cells obtained from *Hhl*/+ mice were *lacZ*-positive (for example, see Figures 3A & B). Both lines demonstrated a coincidence of *lacZ* expression, detected by FDG staining, within cells morphologically resembling megakaryocytes (Figures 3C & D).

This pattern of *lacZ* expression was further supported by the expression of *lacZ* within most megakaryocyte colony forming units (CFU-Mk) generated from *Hhl*/+ and *Hzfl*+ bone marrow in a serum-free agarose assay in the presence of TPO, IL-3 and SLF (Figure 4). In addition to expression in megakaryocytes, *lacZ* was expressed in a population of cells within CFU-M and CFU-GEMM colonies, with no expression observed in CFU-GM, G or BFU-E colonies derived from *Hzfl*+ bone marrow cells (Figure 4). In contrast, *lacZ* was expressed to high levels in BFU-E and cells within CFU-GEMM colonies from *Hhl*/+ bone marrow cells. Expression was also observed in some cells within CFU-M and the macrophage component of CFU-GM (Figure 4); however, granulocytes did not express *lacZ*. Thus, the *in vitro* hematopoietic expression of both Hhl and Hzf was recapitulated *in vivo*.

To define more precisely the cells that expressed *lacZ*, bone marrow cells were selected on the basis of their expression of the megakaryocyte cell surface marker, CD61, by fluorescence-activated cell sorting. Approximately 5% of bone marrow cells derived from *Hzfl*+ mice expressed *lacZ*. About 0.5-1% of the *lacZ*-positive cells did not express CD61. Within the $CD61^+$ fraction obtained from *Hhl*/+ bone marrow cells, approximately 14% of the cells expressed *lacZ* whereas 33% demonstrated *lacZ* expression in the $CD61^-$ fraction.

Sequence analysis of *Hzf* and *Hhl*. To determine the primary sequence of the *Hzf* and *Hhl* trapped genes, *lacZ* fusion transcripts were cloned by 5' rapid amplification of cDNA ends (RACE; Figure 5). Primers corresponding to the sequence obtained by 5'RACE were used for subsequent 3'-RACE to obtain further sequence downstream of the vector integration site (Figure 5 & 6). The full length coding sequence for *Hzf* was obtained by 3' RACE (Figure 6B.). A combination of RACE, inverse PCR and cDNA library screening was used to obtain partial sequences of *Hhl* with a full length coding sequence being obtained using 3' RACE (Figure 6A).

***Hhl*.** The *Hhl* sequence is 1,258 base pairs (bp) in length and contains an uninterrupted open reading frame (ORF) which encodes a putative polypeptide of 298 amino acids (Figure 6A). The sequence of this ORF does not share significant similarity with any known genes; however, the 3' end of *Hhl* shares 88% DNA similarity over 126 bases to a murine expressed sequence tag (EST) isolated from a heart cDNA library (accession # AA919544) using the NCBI search program. Analysis of the translated sequence demonstrated significant homology (33.47%, determined by using Bestfit and GAP programs from GCG) to a peptide encoded by a series of *C.elegans* cDNAs (accession # U41540) using Psi blast from the NCBI search program as well as FASTA3 from the ExPasy Program. The cloned *Hhl* cDNA may not contain the full-length message because it contains no classical AATAAA polyadenylation signal apparent in its 3' UTR. Analysis of the translated sequence using ProfileScan from the ExPasy Program identified a putative phosphotyrosine binding domain between amino acids 78-192.

Hzf. The *Hzf* sequence is 2,300 bp in length and contains a 396 amino acid ORF with a putative translation start site at nucleotide 58 embedded in a consensus Kozak sequence. There is a single AATAAA polyadenylation signal in the 3'UTR. Sequence analysis demonstrated that *Hzf* is a novel gene encoding a polypeptide that includes 3 zinc finger motifs of the C_2H_2 type (Figure 6B). The Hzf putative peptide has sequence similarity (22-32%) with several C_2H_2 type zinc finger containing polypeptides, including the *Xenopus laevis* dsRBP-Zfa (41), the murine p53-inducible zinc finger polypeptide, Wig-1 (42) and its rat homolog PAG608 polypeptide (43).

RNA *in situ* hybridization and northern analysis. To determine whether the expression pattern of *lacZ* described above reflected the endogenous expression pattern of the trapped genes, the X-gal staining pattern in 14.5 d.p.c heterozygous embryos was compared with the pattern of expression observed by RNA *in situ* hybridization analysis in wild type 14.5 d.p.c embryos. In general, the RNA *in situ* signals for both *Hzf* and *Hhl* coincided with the patterns observed by staining for *lacZ* expression (data not shown). Northern blot analysis revealed *Hhl* transcripts of 3 different sizes in various tissues, suggesting that *Hhl* undergoes differential splicing in a tissue-specific manner in adult tissues (Figure 7A). A single 2.3kb *Hzf* message was observed in the bone marrow and brains of adult mice and at lower levels in the thymus (Figure 7B).

Mice homozygous for the *Hzf* and *Hhl* gene trap insertion are viable and fertile. Hzf and Hhl heterozygous mice do not exhibit any apparent abnormalities and are fertile. Intercrosses of F_1 heterozygous mice for each of the Hhl and Hzf lines generated viable homozygous offspring at the appropriate Mendelian ratios. Mice homozygous for either insertion developed normally and did not exhibit any overt phenotype, including the numbers of types of hematopoietic progenitor cells, as determined by *in vitro* colony assays (data not shown). The absence of a discernible phenotype in these mice might reflect the absence of any effect on gene expression as the result of the integration of the PT1 vector into the *Hhl* and *Hzf* genes. To test this possibility, expression of both *Hzf* and *Hhl* RNA transcripts was analyzed in the brains of wild type and homozygous mice by RT-PCR using primers that span the integration site. For both *Hzf* and *Hhl*, cDNA was amplified from both *Hzf* and *Hhl* homozygotes, suggesting that transcripts that extend 3' to the integration site were present (Figure. 8). This conclusion was further confirmed by Northern blot analysis using probes 3' to the site of vector integration. This analysis demonstrated no differences in the levels or size of the transcripts obtained from wild-type or homozygous mice. As shown in Figure 6A, vector integration within the *Hhl* locus was within the 5' UTR upstream of the first coding exon. For *Hzf*, vector integration occurred downstream of the translational start site but upstream of an additional start site which also has a good Kozak consensus; therefore, alternative initiation codon usage may have occurred (Figure 6B). Thus, the sites of vector integration could account for the lack of mutagenesis of these trapped loci. Alternatively, splicing around the gene trap vector may have occurred which would also result in the expression of endogenous transcript. The latter has been reported to occur for a number of trapped genes [44,45,46,47,48,49] and splicing around targeting constructs has also been shown to occur, resulting in the generation of partially functional polypeptides rather than the expected null mutation [50].

DISCUSSION

Novel genes regulated during hematopoietic development were identified and characterized. ES cell clones containing a random gene trap insertion were induced to differentiate into hematopoietic cells by co-culture on OP9 stromal cells. Clones were screened for *lacZ* expression in undifferentiated cells as well as at various stages during hematopoietic differentiation. From a total of 1,350 ES clones, three clones exhibited *lacZ* expression within hematopoietic cells and two of these clones displayed a regulated expression pattern *in vitro* which was recapitulated during hematopoietic development *in vivo*. Molecular cloning and cDNA sequence analysis of the trapped genes in these two ES clones revealed that both genes, which have been denoted as *Hzf* and *Hhl*, are novel. Thus, this *in vitro* expression-based gene trap strategy is a successful approach for screening for novel hematopoietic genes.

In undifferentiated Hzf ES cells, *lacZ* expression was observed in only a few cells and this low frequency of expression was maintained throughout hematopoietic differentiation *in vitro*. The expression of *Hhl* was upregulated being first detected in mesodermal colonies and later expressed in the majority of hematopoietic cells by day 12. During embryogenesis, neither *Hzf* or *Hhl* was expressed in hematopoietic cells within blood islands in the extraembryonic yolk sac, the site of primitive hematopoiesis. Neither gene was expressed within the intraembryonic para-aortic splanchnopleur/aorta-gonad-mesonephros (p-Sp/AGM) region, the site where definitive hematopoiesis is thought to be initiated [51,52]. This lack of detectable *Hhl* or *Hzf* expression in these hematopoietic compartments does not exclude the possibility that cells within these regions express these genes. It is possible that in these compartments only a low frequency of cells express *Hzf* or *Hhl* and/or expression in cells within these regions is below the level of detection.. Expression of both trapped genes in hematopoietic cells was observed first in the fetal liver and subsequently adult bone marrow. The low frequency of *Hzf* expression observed *in vitro* was recapitulated *in vivo* in the embryonic liver and in only 2 % of adult bone marrow cells. The higher frequency of *Hhl* expression within day 12 hematopoietic cells *in vitro* was also recapitulated *in vivo* where 80% of adult bone marrow cells expressed *lacZ*. The majority of cells expressing *Hzf* were megakaryocytes, as determined by morphology, *lacZ* expression within CFU-Mk, and cell surface expression of the megakaryocyte-specific surface marker, CD61. *Hzf* was also expressed in other hematopoietic cell lineages contained within CFU-GEMM and at a very low frequency in CFU-M. *Hhl* was more widely expressed than *Hzf* within hematopoietic lineages, particularly in erythroid and megakaryocyte cells as well as in the majority of cells comprising CFU-GEMM.

Both *Hzf* and *Hhl* were expressed in megakaryocytes, suggesting that these novel genes may play a role in the differentiation and/or maturation of this cell lineage. The major determinants of commitment to the megakaryocyte lineage and the regulation of gene expression in this lineage remain unclear. The generation of several knock-out mice have aided in our understanding of megakaryocyte differentiation. Mice lacking thrombopoietin (TPO) or its receptor c-Mpl have reduced numbers of megakaryocytes and as a consequence are also thrombocytopenic [8,53]. These mice also have reduced numbers of hematopoietic progenitors of multiple lineages [54,55]. Mice with a megakaryocyte-selective loss of GATA-1 and NF-E2 null mice are also thrombocytopenic as a result of arrested megakaryocyte differentiation and maturation [56,57]. Mice lacking GATA-1 or NF-E2 also display defects in

erythropoiesis [58,57], suggesting that these transcription factors are co-expressed within a bipotential stem cell capable of giving rise to both megakaryocytes and erythroid cells. Recently, it has been shown that the multiple zinc finger polypeptide, Friend of GATA-1 (FOG), cooperates with GATA-1 to promote erythroid and megakaryocyte differentiation [59]. In mice lacking FOG due to a targeted null mutation in ES cells, the differentiation of erythroid precursors is blocked [60]. However, in contrast to GATA-1 deficient mice, megakaryocytes are completely absent, suggesting a GATA-1 independent role for FOG perhaps at the very earliest stage of megakaryocyte differentiation from a bipotential erythroid/megakaryocyte progenitor. In this study, *Hzf* expression was observed in megakaryocytes but not in erythroid cells, suggesting that the function of this novel gene may be restricted to the megakaryocyte lineage. The presence of three zinc-finger motifs in Hzf suggests that it may also be involved in transcriptional regulation in megakaryocytes. In contrast, *Hhl* was strongly expressed in both erythroid and megakaryocyte lineages. Given that both *Hzf* and *Hhl* were also expressed within multipotential CFU-GEMM colonies, these novel genes may be expressed in a bi/multipotential progenitor.

In addition to expression in hematopoietic cells, *Hhl* was also expressed in the embryonic and adult heart and *Hhl* has significant homology to an EST isolated from a heart cDNA library. Northern blot analysis of adult mouse tissue also demonstrated *Hhl* transcripts in the brain and kidney. In adult tissue, *Hzf* was also expressed in the brain. Thus, these genes may also be involved in other developmental processes outside of the hematopoietic system.

The gene trap strategy resulted in the identification of two novel genes with coincidence of expression within megakaryocytes. Insertion of the gene trap vector into *Hzf* and *Hhl* has allowed the molecular cloning of both genes and has also facilitated the rapid analysis of their expression during mouse embryogenesis.

Example 2

Mice with true null alleles for Hzf were generated. A targeting vector was designed to replace a 5.5kb genomic fragment containing exons encoding 3 zinc finger domains with a IRES LacZ and a neomycin resistance gene. The IRES LacZ and Neo resistance gene was inserted in the targeting vector in the sense orientation to the Hzf transcript. The targeting vector was electroporated into R1 ES cells. After positive and negative selection, 6 independent clones were screened after checking for homologous recombination using 5 prime and 3 prime flanking probes. A Neo probe was also used to check single integration. Three out of 6 clones were aggregated, resulting in germline transmission. Heterozygous offspring were interbred to generate homozygotes. The total number of new born pups decreased at 3 weeks after birth. In order to determine the stage of neonatal development affected by the Hzf mutation, neonatal genotyping was performed. One week old Hzf -/- pups were viable, usually of normal size and occurred at the expected Mendelian frequency. However, thereafter, typically from 2 to 3 weeks after birth, the percentage of homozygotes was decreased. Most of the Hzf -/- showed behavioral abnormalities. Six out of 24 Hzf -/- mice were able to lift their heads only to the left side. Seventeen out of 24 exhibited tremoring of their bodies when they were put on unstable surfaces (e.g. lid of the cage). Surviving mice

were smaller than their littermates. Hemorrhage was frequently observed (e.g. at 3 weeks) in the brain of Hzf homozygous mutants. Therefore, in summary the following phenotype was observed:

(a) The frequency of viable Hzf -/- mice decreased at the age of two weeks, although normal Mendelian ratios of Hzf mutants was observed at one week after birth.

(b) Hemorrhage was frequently observed in the brain of Hzf -/- mice.

(c) The majority of surviving Hzf mice were smaller than their littermates, presumably as a result of sustained hemorrhage during a period of rapid growth; they were able to lift their heads only to the right or left; and their bodies tremor when placed on unstable surfaces.

The phenotype resembled the phenotype in the case of sustained brain damage resulting from a stroke in humans.

The ultrastructure of mature megakaryocytes from the Hzf -/- mouse was compared to mature megakaryocytes from the bone marrow of a Hzf +/+ mouse. The cytoplasm of megakaryocytes from the Hzf +/+ mouse showed platelet fields or territories that are clearly demarcated and many granules were observed. In contrast, the cytoplasm of megakaryocytes of the Hzf -/- mouse showed that the organization of the demarcation membrane system is aberrant, and granules are sparse.

The present invention is not to be limited in scope by the specific embodiments described herein, since such embodiments are intended as but single illustrations of one aspect of the invention and any functionally equivalent embodiments are within the scope of this invention. Indeed, various modifications of the invention in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description and accompanying drawings. Such modifications are intended to fall within the scope of the appended claims.

All publications, patents and patent applications referred to herein are incorporated by reference in their entirety to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated by reference in its entirety. All publications, patents and patent applications mentioned herein are incorporated herein by reference for the purpose of describing and disclosing the cell lines, vectors, methodologies etc. which are reported therein which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

It must be noted that as used herein and in the appended claims, the singular forms "a", "an", and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, reference to "a host cell" includes a plurality of such host cells, reference to the "antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Table 1
Summary of *lacZ* expression within ES cell gene trap clones upon hematopoietic differentiation[a]

Number of clones	Days after induction			
	0	5	10	
White clones				
1278	−	−	−	
1	−	+	−	7H7
1	−	+	+	Hhl
Blue clones				
2	+	+	+	4A11 and Hzf
24	+	+	−	
44	+	−	−	
1350 total				

[a] White clones, *lacZ* negative in undifferentiated ES cells; blue clones, *lacZ* positive in undifferentiated ES cells; +, *lacZ* positive during differentiation; −, *lacZ* negative during differentiation.

Referances:

1. Bernstein A, Forrester L, Reith AD, Dubreuil P, Rottapel R: The murine *W/c-kit* and *Steel* loci and the control of hematopoiesis. Semin. Hematol. 28:138-142, 1991
2. Bedell MA, Jenkins NA, Copeland NG: Mouse models of human disease. Part I: Techniques and resources for genetic analysis of mice. Genes & Dev. 11:1-10, 1997
3. Bedell MA, Largaespada DA, Jenkins NA, Copeland NG: Mouse models of human disease. Part II: Recent progress and future directions. Genes & Dev. 11: 11-43, 1997
4. Shivdasani RA and Orkin SH: The transcriptional control of hematopoiesis. Blood 87: 4025, 1996
5. Bernstein A, Correll P, Paulson R, Potter M: *In vitro* and *in vivo* approaches to the molecular analysis of mouse hematopoiesis. In Mammalian Development ed Peter Lonai, pg 101, 1996
6. Fischer, K-D Zmuidzinas A, Gardner S, Barbacid M, Bernstein A Guidos C. Defective T-cell receptor signalling and positive selection of *vav*-deficient CD4+ CD8+ thymocytes. Nature 374:474, 1995
7. Shalaby F, Rossant J, Yamaguchi TP, Gertstenstein M, Wu X-F, Breitman ML, Shuh AC: Failure of blood-island formation and vasculogenesis in *Flk*-1-deficient mice. Nature 376:62, 1995
8. Gurney AL., Carver-Moore K, de Sauvage FJ, Moore MW: Thrombocytopenia in c-mpl-deficient mice. Science 265:1445, 1994
9. Cheng J, Baumhueter S, Cacalano, G Thibodeaux H, Thomas R, Broxmeyer HE, Cooper S, Hague N, Moore M, Lasky LA: Hematopoietic defects in mice lacking the sialomucin CD34. Blood 87:479, 1996.
10. Lieschke GJ, Grail D, Hodgson G, Metcalf D, Stanley E, Cheers C, Fowler KJ, Basu S, Zhan YK, Dunn AR: Mice lacking granulocyte colony-stimulating factor have chronic neutropenia, granulocyte and macrophage progenitor cell deficiency, and impaired neutrophil mobilization. Blood 84: 1737, 1994
11. Gossler A, Joyner AL, Rossant J, Skarnes WC: Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. Science 244: 463, 1989
12. Friedrich G and Soriano P: Promoter traps in embryonic stem cells: A genetic screen to identify and mutate developmental genes in mice. Genes & Dev, 5: 1513, 1991
13. Skarnes WC, Auerbach BA, Joyner AL: A gene trap approach in mouse embryonic stem cells: the lacZ reporter is activated by splicing, reflects endogenous gene expression, and is mutagenic in mice. Genes & Dev. 6: 903, 1992
14. von Melchner H, DeGregori JV, Rayburn H, Reddy S, Friedel C, Ruley HE. Selective disruption of genes expressed in totipotent embryonal stem cells. Genes & Dev. 6: 919, 1992
15. Wurst W, Rossant J, Prideaux V, Kowanacka M, Joyner AL, Hill DP, Guillemot F, Gasca S, Cado D, Auerbach A, Ang S-L. A large-scale gene-trap screen for insertional mutations in developmentally regulated genes in mice. Genetics 139:889, 1995
16. Evans MJ, Carlton MBL, Russ AP. Gene trapping and functional genomics. Trends Genet. 13:370, 1997
17. Sam M, Wurst W, Klüppel M, Jin O, Heng H, Bernstein A: *Aquarius*, a novel gene isolated by gene

trapping with an RNA-dependent RNA polymerase motif. Dev. Dyn. 212: 304, 1998

18. Stanford WL, Caruana G, Vallis KA, Inamdar M, Hidaka M, Bautch VL, Bernstein A: Expression trappping: Identification of novel genes expressed in hematopoietic and endothelial lineages by gene trapping in ES cells. Blood 92: 4622, 1998.
19. Doetschman TC, Eistetter H, Katz M, Schmidt W, Kemler R: The in vitro development of blastocyst-derived embryonic stem cell lines: Formation of visceral yolk sac, blood islands and myocardium. J. Embryol. Exp. Morphol. 87: 27, 1985
20. Bain G, Kitchens D, Yao M, Huettner JE, Gottlieb DI: Embryonic stem cells express neuronal properties *in vitro*. Dev. Biol.168:342, 1995
21. Maltsev VA, Rohwedel J, Hescheler J, Wobus AM: Embryonic stem cells *differentiate in vitro* into cardiomyocytes representing sinusnodal, atrial and ventricular cell types. Mech. Dev. 44: 41, 1993
22. Wang R, Clark R, Bautch VL: Embryonic stem cell-derived cystic embryoid bodies form vascular channels: an *in vitro* model of blood vessel development. Development 114:303, 1992
23. Rohwedel J, Maltsev V, Bober E, Arnold H-H, Hescheler J, Wobus AM: Muscle cell differentiation of embryonic stem cells reflects myogenesis *in vivo*: Developmentally regulated expression of myogenic determination genes and functional expression of ionic currents. Dev. Biol. 164:87, 1994.
24. Keller G, Kennedy M, Papayannopoulou T, Wiles MV: Hematopoietic commitment during embryonic stem cell differentiation in culture. Mol. Cell. Biol. 13: 473, 1993
25. Wiles MV and Keller G: Multiple hematopoietic lineages develop from embryonic stem (ES) cells in culture. Development 111: 259, 1991
26. Schmitt RM, Bruyns E, Snodgrass R: Hematopoietic development of embryonic stem cells in vitro: cytokine and receptor gene expression. Genes & Development 5: 728, 1991
27. Palacios R, Bucana C, Xie X: Long-term culture of lymphohematopoietic stem cells. Proc. Natl. Acad. Sci. USA 93: 5247, 1996
28. Kennedy M, Firpo M, Choi K, Wall C, Robertson S, Kabrun N, Keller G: A common precursor for primitive erythropoiesis and definitive haematopoiesis. Nature 386: 488, 1997
29. Forrester LM, Nagy A, Sam M, Watt A, Stevenson L, Bernstein A, Joyner AL, Wurst W: An induction gene trap screen in embryoid stem cells: Identification of genes that respond to retinoic acid in vitro. Proc. Natl. Acad. Sci. USA 93:1677, 1996
30. Shirai M, Miyashita A, Ishii N, Itoh Y, Satokata I, Watanabe YG, Kuwano R: A gene trap strategy for identifying genes expressed in the embryonic nervous system. Zool. Sci. 13:277, 1996
31. Baker RK, Haendel MA, Swanson BJ, Shambaugh JC, Micales BK, Lyons GE: *In vitro* preselection of gene-trapped embryonic stem cell clones for characterizing novel developmentally regulated genes in the mouse. Dev. Biol. 185:201, 1997
32. Muth K, Bruyns R, Thorey IS, von Melchner H: Disruption of genes regulated during hematopoietic differentiation of mouse embryonic stem cells. Dev Dyn. 212: 277, 1998
33. Nakano T, Kodama H, Honjo T: Generation of lymphohematopoietic cells from embryonic stem cells in culture. Science 265: 1098, 1994

34. Nagy ARJ, Rossant J, Nagy R, Abramow-Newerly W, Roder JC: Derivation of completely cell culture-derived mice from early passage embryonic stem cells. Proc. Natl. Acad. Sci. USA. 90: 8424, 1993

35. Nagy ARJ: In Gene Targeting: A Practical Approach, J.A.L., ed (Oxford: IRL), pp. 147-178 (1993)

36. Hill D and Wurst W: Screening for novel pattern formation genes using gene trap approaches. Methods Enzymol. 225: 664, 1993

37. von Melchner H, Reddy S, Ruley HE. Isolation of cellular promoters by using a retrovirus promoter trap. Proc Natl Acad Sci U S A. 87: 3733, 1990

38. Sambrook J, Fritsch EF, Maniatis T: Molecular Cloning. A Laboratory Manual. Cold Spring Harbor: Cold Spring Harbor Laboratory Press, 1989

39. Fiering SN, Roederer M, Nolan GP, Micklem DR, Parks DR, Herzenberg LA: Improved FACS-Gal: Flow cytometric analysis and sorting of viable eukaryotic cells expressing reporter gene constructs. Cytometry 12:291, 1991

40. Hogge D, Fanning S, Bockhold K, Petzer A, Lambie K, Lansdorp P, Eaves A, Eaves C: Quantitation and characterization of human megakaryocyte colony-forming cells using a standardized serum-free agarose assay. Brit. J. Haem. 96: 790, 1997.

41. Finerty PJ and Bass BL: A Xenopus zinc finger polypeptide that specifically binds dsRNA and RNA-DNA hybrids. J. Mol. Biol. 271: 195, 1997

42. Varmeh-Ziaie S, Okan I, Wang Y, Magnusson KP, Warthoe P, Strauss M, Wiman KG: Wig-1, a new p53-induced gene encoding a zinc finger polypeptide. Oncogene 15: 2699, 1997

43. Israeli D, Tessler E, Haupt Y, Elkeles A, Wilder S, Amson R, Telerman A, Oren M: A novel p53-inducible gene, PAG608, encodes a nuclear zinc finger polypeptide whose overexpression promotes apoptosis. EMBO. 16:4384, 1997

44. Gasca S, Hill DP, Klingensmith J, Rossant J: Characterization of a gene trap insertion into a novel gene, cordon-bleu, expressed in axial structures of the gastrulating mouse embryo. Dev. Genetics 17: 141, 1995

45. Serafini T, Colamarino SA, Leonardo ED, Wang H, Beddington R, Skarnes WC, Tessier-Lavigne M: Netrin-1 is required for commissural axon guidance in the developing vetebrate nervous system. Cell 87: 1001, 199

46. Yeo TT, Yang T, Massa SM, Zhang JS, Honkaniemi J, Butcher LL, Longo FM: Deficient LAR expression decreases basal forebrain cholinergic neuronal size and hippocampal cholinergic innervation. J. Neurosci. Res. 47: 348, 1997

47. Faisst, AM and Gruss P: *Bodenin*: A novel murine gene expressed in restricted areas of the brain. Dev. Dyn. 212; 293, 1998

48. Voss AK, Thomas T, Gruss P: Compensation for a gene trap mutation in the murine microtubule-associated polypeptide 4 locus by alternative polyadenylation and alternative splicing. Dev Dyn. 212: 258, 1998

49. McClive P, Pall G, Newton K, Lee M, Mullins J, Forrester L: Gene trap integrations expressed in the developing heart: Insertion site affects splicing of the PT1-ATG vector. Dev Dyn. 212: 267, 1998

50. Melet F, Motro B, Rossi DJ, Zhang L, Bernstein A: Generation of a novel Fli-1 polypeptide by gene targeting leads to a defect in thymus development and a delay in Friend virus-induced erythroleukemia. Mol. Cell. Biol. 16: 2708, 1996

51. Dzierzak E, and Medvinsky A: Mouse embryonic hematopoiesis. Trends Genet. 11: 359, 1995

52. Medvinsky A and Dzierzak E: Definitive hematopoiesis is autonomously initiated by the AGM region. Cell. 86: 897, 1996

53. deSauvage FJ, Carver-Moore K, Luoh S-M, Ryan A, Dowd M, Eaton DL, Moore MW. Physiological regulation of early and late stages of megakaryocytopoiesis by thrombopoietin. J. Exp Med 183:651, 1996

54. Alexander WS, Roberts AR, Nicola NA, Li R, Metcalf D: Deficiencies in progenitor cells of multiple hematopoietic lineages and defective megakaryocytopoiesis in mice lacking the thrombopoietin receptor c-Mpl. Blood 87:2162, 1996

55. Carver-Moore K, Broxmeyer HE, Luoh S-M, Ryan A, Dowd M, Eaton DL. Moore MW, deSauvage FJ: Low levels of erythroid and myeloid progenitors in thrombopoietin and c-*mpl*-deficient mice. Blood 88:803, 1996

56. Shivdasani RA, Fukiwara Y, McDevitt MA, Orkin SH: A lineage-selective knockout establishes the critical role of transcription factor GATA-1 in megakaryocyte growth and platelet development. EMBO 16: 3965, 1997

57. Shivdasani RA, Rosenblatt MF, Zucker-Franklin D, Jackson CW, Hunt P, Saris CJ, Orkin SH: Transcription factor NF-E2 is required for platelet formation independent of the actions of thrombopoietin/MGDF in megakaryocyte development. Cell 81:695, 1995

58. Fujiwara Y, Browne CP, Cunniff K, Goff SC, Orkin SH: Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. PNAS 93: 12355, 1996

59. Tsang AP, Visvader JE, Turner CA, Fujiwara Y, Yu C, Weiss MJ, Crossley M, Orkin SH: FOG, a multiple zinc finger polypeptide, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation. Cell 90: 109, 1997

60. Tsang AP, Fujiwara Y, Hom DB, Orkin SH: Failure of megakaryopoiesis and arrested erythropoiesis in mice lacking the GATA-1 transcriptional cofactor FOG. Genes and Dev 12: 1176, 1998

WE CLAIM:

1. An isolated Hzf nucleic acid molecule of at least 30 nucleotides which hybridizes to SEQ ID NO. 1, or the complement of SEQ ID NO. 1, under stringent hybridization conditions.
2. An isolated Hhl nucleic acid molecule of at least 30 nucleotides which hybridizes to SEQ ID NO. 3 or the complement of SEQ ID NO.3, under stringent hybridization conditions.
3. An isolated nucleic acid molecule which comprises:
   (i) a nucleic acid sequence encoding a polypeptide having substantial sequence identity with the amino acid sequence of SEQ. ID. NO.2 or SEQ. ID. NO 4.;
   (ii) nucleic acid sequences complementary to (i);
   (iii) a degenerate form of a nucleic acid sequence of (i);
   (iv) a nucleic acid sequence comprising at least 18 nucleotides and capable of hybridizing to a nucleic acid sequence in (i), (ii), or (iii);
   (v) a nucleic acid sequence encoding a truncation, an analog, an allelic or species variation of a polypeptide comprising the amino acid sequence of SEQ. ID. NO.2, or SEQ. ID. NO 4; or
   (vi) a fragment, or allelic or species variation of (i), (ii) or (iii).
4. An isolated nucleic acid molecule which comprises:
   (a) a nucleic acid sequence having substantial sequence identity or sequence similarity with a nucleic acid sequence of SEQ. ID. NO. 1 or 3;
   (b) nucleic acid sequences complementary to (i), preferably complementary to the full nucleic acid sequence of SEQ. ID. NO. 1 or 3;
   (c) nucleic acid sequences differing from any of the nucleic acid sequences of (i) or (ii) in codon sequences due to the degeneracy of the genetic code; or
   (d) a fragment, or allelic or species variation of (i), (ii) or (iii).
5. An isolated nucleic acid molecule which encodes a protein which binds an antibody of a Hzf or Hhl polypeptide.
6. A regulatory sequence of an isolated nucleic acid molecule as claimed in any of the preceding claims fused to a nucleic acid which encodes a heterologous protein.
7. A vector comprising a nucleic acid molecule of any of the preceding claims.
8. A host cell comprising a nucleic acid molecule of any of the preceding claims.
9. An isolated Hzf polypeptide comprising an amino acid sequence of SEQ. ID. NO. 2.
10. An isolated Hhl polypeptide comprising an amino acid sequence of SEQ. ID. NO. 4.
11. An isolated polypeptide having at least 50% amino acid sequence identity to an amino acid sequence of SEQ. ID. NO. 2 or 4.
12. A method for preparing a polypeptide as claimed in claim 9 or 10 comprising:
   (a) transferring a vector as claimed in claim 7 into a host cell;
   (b) selecting transformed host cells from untransformed host cells;

(c) culturing a selected transformed host cell under conditions which allow expression of the protein; and

(d) isolating the polypeptide.

13. A polypeptide prepared in accordance with the method of claim 12.

14. An antibody having specificity against an epitope of a polypeptide as claimed in claim 9 or 10.

15. An antibody as claimed in claim 14 labeled with a detectable substance and used to detect the polypeptide in biological samples, tissues, and cells.

16. A probe comprising a sequence encoding a polypeptide as claimed in claim 9 or 10, or a part thereof.

17. A method of diagnosing and monitoring conditions mediated by a polypeptide as claimed in claim 9 or 10 by determining the presence of a nucleic acid molecule as claimed in any of the preceding claims or a polypeptide as claimed in any of the preceding claims.

18. A method as claimed in claim 17 wherein the condition is a disorder of the hematopoietic system.

19. A method for identifying a substance which associates with a polypeptide as claimed in claim 9 or 10 comprising (a) reacting the polypeptide with at least one substance which potentially can associate with the polypeptide, under conditions which permit the association between the substance and polypeptide, and (b) removing or detecting polypeptide associated with the substance, wherein detection of associated polypeptide and substance indicates the substance associates with the polypeptide.

20. A method for evaluating a compound for its ability to modulate the biological activity of a polypeptide as claimed in claim 9 or 10 comprising providing a known concentration of the polypeptide with a substance which associates with the polypeptide and a test compound under conditions which permit the formation of complexes between the substance and polypeptide, and removing and/or detecting complexes.

21. A method for detecting a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence of SEQ. ID. NO. 2 or 4 in a biological sample comprising the steps of:

(a) hybridizing a nucleic acid molecule of claim 4 to nucleic acids of the biological sample, thereby forming a hybridization complex; and

(b) detecting the hybridization complex wherein the presence of the hybridization complex correlates with the presence of a nucleic acid molecule encoding the polypeptide in the biological sample.

22. A method as claimed in claim 21 wherein nucleic acids of the biological sample are amplified by the polymerase chain reaction prior to the hybridizing step.

23. A method for treating a condition mediated by a polypeptide as claimed in claim 9 or 10 comprising administering an effective amount of an antibody as claimed in claim 14 or a substance or compound identified in accordance with a method claimed in claim 19 or claim 20.

24. A method as claimed in claim 23 wherein the condition is a disorder of the hematopoietic system.

25. A composition comprising one or more of a nucleic acid molecule or polypeptide claimed in any of the preceding claims, or a substance or compound identified using a method as claimed in any of the preceding claims, and a pharmaceutically acceptable carrier, excipient or diluent.

26. Use of one or more of a nucleic acid molecule or polypeptide claimed in any of the preceding claims, or a substance or compound identified using a method as claimed in any of the preceding claims in the preparation of a pharmaceutical composition for treating a condition mediated by a polypeptide as claimed in claim 9 or 10.

27. A transgenic non-human mammal wherein a nucleic acid molecule as claimed in claim 3, or a portion thereof is deleted in the mammal's germ cells and somatic cells.

28. A transgenic non-human mammal all of whose germ cells and somatic cells contain a recombinant expression vector that inactivates or alters a gene encoding a Hzf polypeptide as claimed in claim 4 resulting in a Hzf associated pathology.

29. A transgenic non-human mammal which doe not express a Hzf polypeptide as claimed in claim 9 resulting in a Hzf associated pathology.

30. A transgenic animal assay system which provides a model system for testing for an agent that reduces or inhibits an Hzf associated pathology

(a) administering the agent to a transgenic non-human animal as claimed in claim 29; and

(b) determining whether said agent reduces or inhibits an Hzf associated pathology in the transgenic mouse relative to a transgenic mouse of step (a) which has not been administered the agent.

FIGURE 1

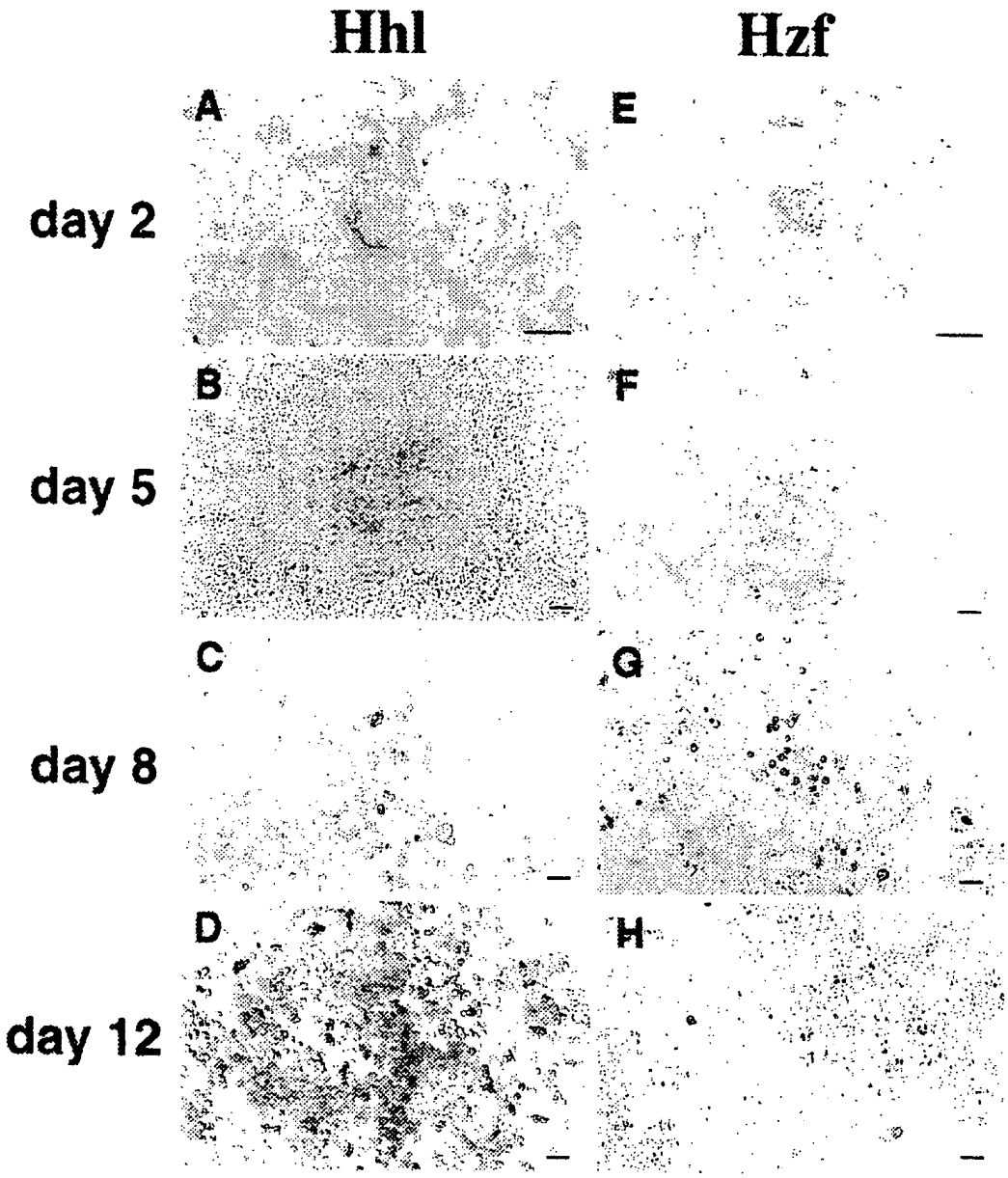


FIGURE 2A-H

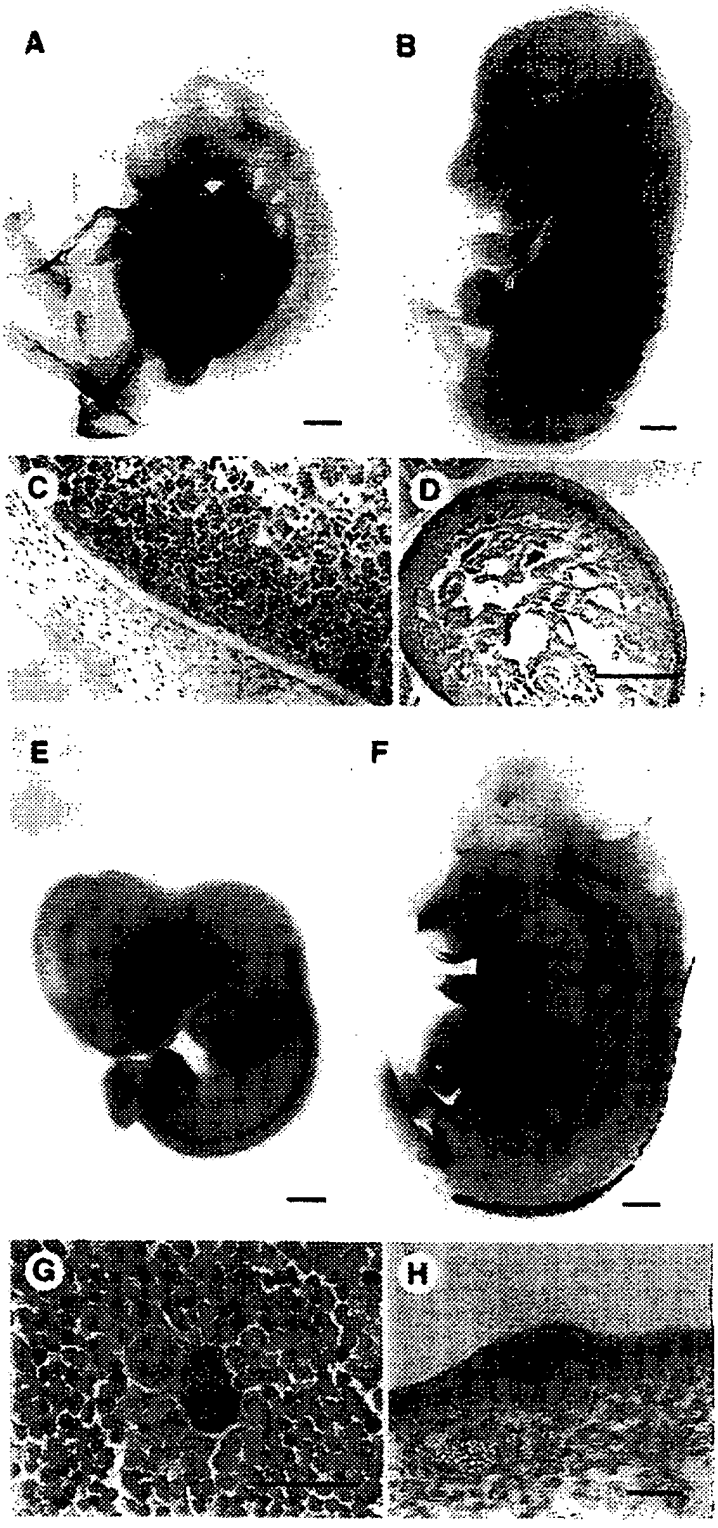


FIGURE 3A-D

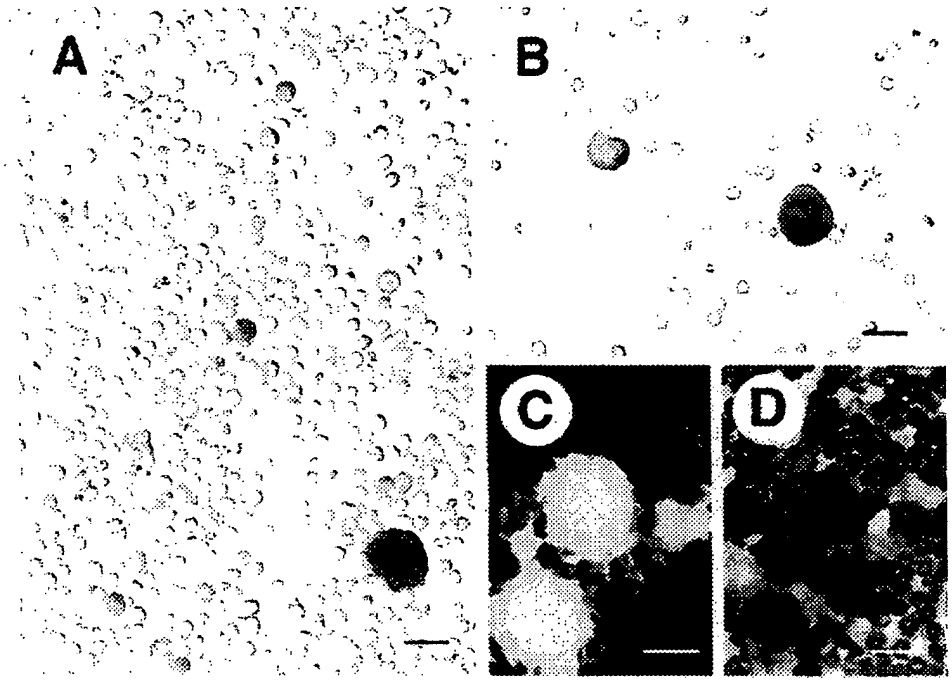


FIGURE 4

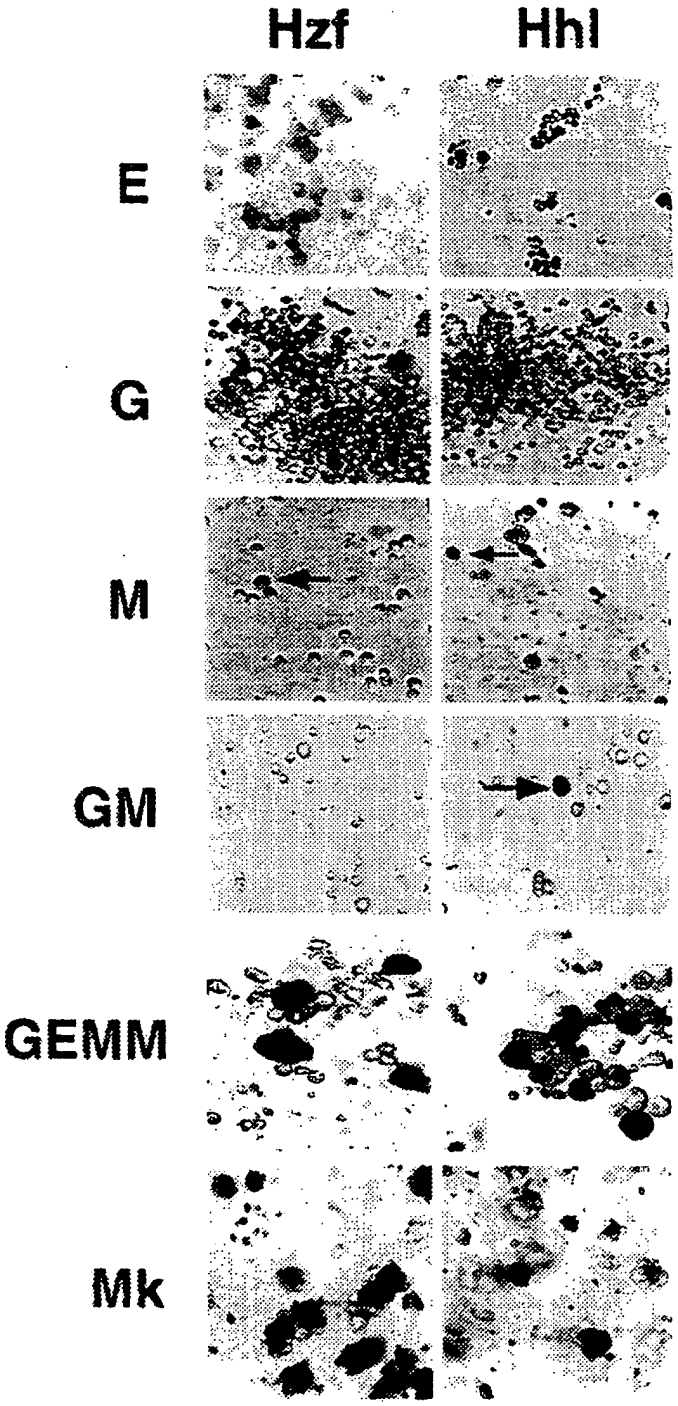


FIGURE 5A

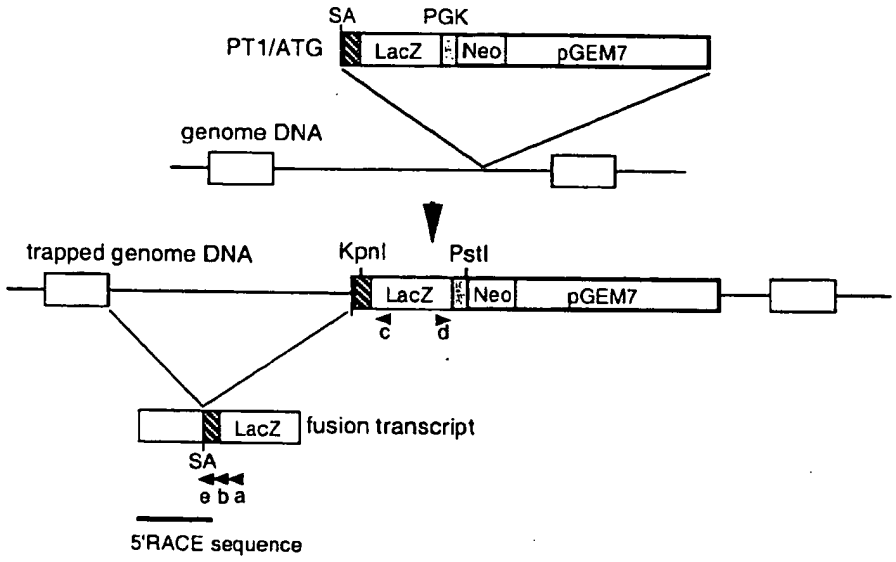

SA
PGK
PT1/ATG
LacZ
Neo
pGEM7
genome DNA
trapped genome DNA
KpnI
PstI
LacZ
Neo
pGEM7
c
d
LacZ
fusion transcript
SA
e b a
5'RACE sequence

FIGURE 5B

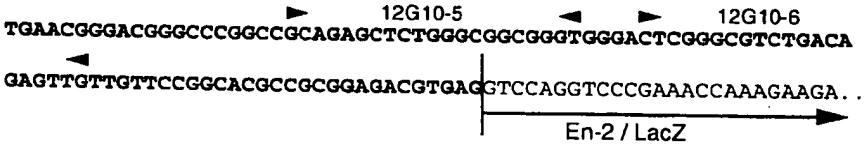

12G10-5
12G10-6
TGAACGGGACGGGCCCGGCCGCAGAGCTCTGGGCGGCGGGTGGGACTCGGGCGTCTGACA
GAGTTGTTGTTCCGGCACGCCGCGGAGACGTGAGGTCCAGGTCCCGAAACCAAAGAAGA..
En-2 / LacZ

FIGURE 5C

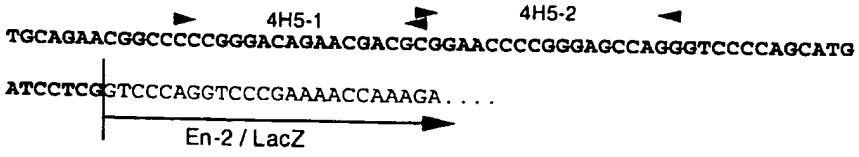

4H5-1
4H5-2
TGCAGAACGGCCCCCGGGACAGAACGACGCGGAACCCCGGGAGCCAGGGTCCCCAGCATG
ATCCTCGGTCCCAGGTCCCGAAAACCAAAGA....
En-2 / LacZ

FIGURE 6A

```
CTCGGGCGTCTGACAGAGTTGTTGTTCCGGCACGCCGCGGAGACGTGAGTTGTTCCAGGT   60
GGATGTGGGCAGAGAGGTTTGCAGAACTGAAATGGAGGTCGGAGCTTCGTTCCAGAAGGT  120
TAGTGGGTCATCTGATTCTGTGGCCACACTGAACAGTGAAGAATTTGTTTTGGTTTCTCA  180
GCACACAGATGCCACTTCTATAAAGGATGATGGGAAGCCACAGCTGAAGATAGCTTCCAA  240
TGGTGATGAGCAGTTGGAAAAAGCCATGGAAGAGATTTTGAGAGATTCCGAGAAAGGACA  300
                          M  E  E  I  L  R  D  S  E  K  G  Q

AAGCGGTCTACCTGTTGATTGCCAAGGATCCAGTGAGATTTCAGACTGTCCTTTTGGAGA  360
 S  G  L  P  V  D  C  Q  G  S  S  E  I  S  D  C  P  F  G  D

TGTGCCGGCCAGCCAAACAACTAAGCCGCCTCTCCAGTTAATTTTGGATCCATCTAATAC  420
 V  P  A  S  Q  T  T  K  P  P  L  Q  L  I  L  D  P  S  N  T

AGAAATTTCCACACCCAGACCATCTTCTCCAAGCAGATTTCCTGAAGAAGACAGTGTTCT  480
 E  I  S  T  P  R  P  S  S  P  S  R  F  P  E  E  D  S  V  L

CTTTAACAAGCTGACATACTTAGGATGTATGAAGGTTTCTTCCCACGCAGTGAAGTGGAG  540
 F  N  K  L  T  Y  L  G  C  M  K  V  S  S  H  A  V  K  W  R

GCTTTACGGCCATGCCACCATGAGAGCTTCCAGTCAGTACCCCTTTGCTGTTACTCTGTA  600
 L  Y  G  H  A  T  M  R  A  S  S  Q  Y  P  F  A  V  T  L  Y

TGTGCCCAATGTTCCAGAAGGATCTGTGAGAATCATAGACCAGTCAAGCAATGTGGAGAT  660
 V  P  N  V  P  E  G  S  V  R  I  I  D  Q  S  S  N  V  E  I

AGCATCTTTTCCAATTTATAAAGTGCTTTTCTGTGCACGTGGGCATGATGAGACAGCCGA  720
 A  S  F  P  I  Y  K  V  L  F  C  A  R  G  H  D  E  T  A  E

GAGCAATTGCTTTGCATTTACAGAGAGTTCTCATGGCTCAGAAGAATTTCAGATACATGT  780
 S  N  C  F  A  F  T  E  S  S  H  G  S  E  E  F  Q  I  H  V

TTTCTCCTGTGAAATTAAAGAGGCAGTAAGCAGAATTTTATATAGTTTCTGCACTGCATT  840
 F  S  C  E  I  K  E  A  V  S  R  I  L  Y  S  F  C  T  A  F

CAAACGTTCTTCCAGACAAGTGTCTGATGTTAAAGACTCAGTCATCCCGACTCCCGACAG  900
 K  R  S  S  R  Q  V  S  D  V  K  D  S  V  I  P  T  P  D  S

TGATGTGTTTACCTTCAGTGTCTCCTTGGAGGTCAAAGAAGATGATGGAAAAGGAAACTT  960
 D  V  F  T  F  S  V  S  L  E  V  K  E  D  D  G  K  G  N  F

TAGTCCCGTGCCTAAGGATAGAGATAAATTTTATATCAAAATAAAGCAAGGAATAGAGAA 1020
 S  P  V  P  K  D  R  D  K  F  Y  I  K  I  K  Q  G  I  E  K

GAAGGTTGTGATTACAGTTCAGCAACTGTCTAACAAAGAATTAGCTATTGAGAGATGTTT 1080
 K  V  V  I  T  V  Q  Q  L  S  N  K  E  L  A  I  E  R  C  F

TGGAATGTTATTAAGCCCAGGTCGAAACGTGAAGAACAGTGACATGCATTTACTGGATAT 1140
 G  M  L  L  S  P  G  R  N  V  K  N  S  D  M  H  L  L  D  M

GGAGTCCCGGAAAGAGCTATGATGGGAGAGCTTACGTCATCACGGGCATGTGGAACCCCA 1200
 E  S  R  K  E  L  *

ACGCACCAATATTTCTGGCTCTTAATGAAGAAACCCTGTCTCAAAAAAAACCAAAAAA   1258
```

FIGURE 6B

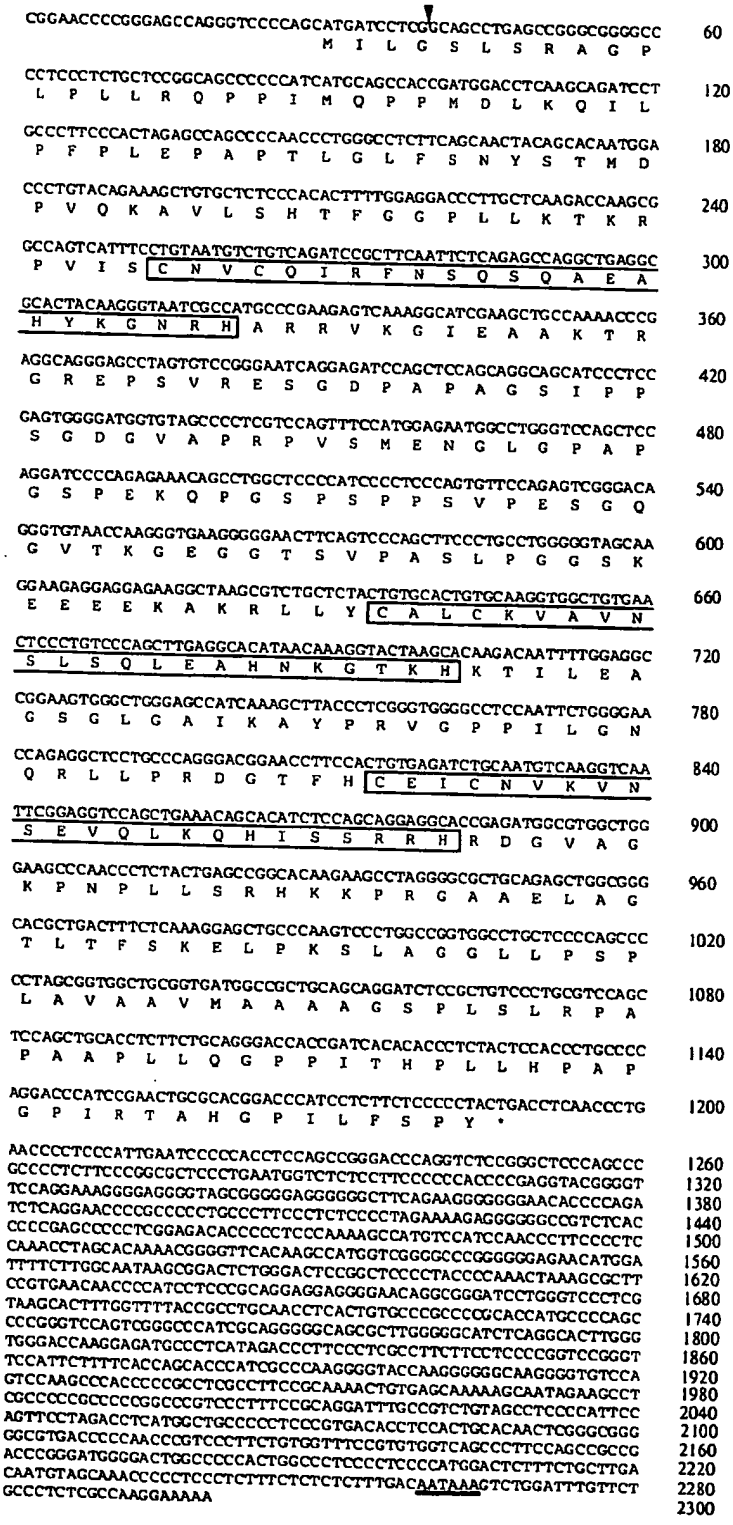

CGGAACCCCGGGAGCCAGGGTCCCCAGCATGATCCTCGGCAGCCTGAGCCGGGCGGGGCC 60
M I L G S L S R A G P
CCTCCCTCTGCTCCGGCAGCCCCCCATCATGCAGCCACCGATGGACCTCAAGCAGATCCT 120
L P L L R Q P P I M Q P P M D L K Q I L
GCCCTTCCCACTAGAGCCAGCCCCAACCCTGGGCCTCTTCAGCAACTACAGCACAATGGA 180
P F P L E P A P T L G L F S N Y S T M D
CCCTGTACAGAAAGCTGTGCTCTCCCACACTTTTGGAGGACCCTTGCTCAAGACCAAGCG 240
P V Q K A V L S H T F G G P L L K T K R
GCCAGTCATTTCCTGTAATGTCTGTCAGATCCGCTTCAATTCTCAGAGCCAGGCTGAGGC 300
P V I S C N V C Q I R F N S Q S Q A E A
GCACTACAAGGGTAATCGCCATGCCCGAAGAGTCAAAGGCATCGAAGCTGCCAAAACCCG 360
H Y K G N R H A R R V K G I E A A K T R
AGGCAGGGAGCCTAGTGTCCGGGAATCAGGAGATCCAGCTCCAGCAGGCAGCATCCCTCC 420
G R E P S V R E S G D P A P A G S I P P
GAGTGGGGATGGTGTAGCCCCTCGTCCAGTTTCCATGGAGAATGGCCTGGGTCCAGCTCC 480
S G D G V A P R P V S M E N G L G P A P
AGGATCCCCAGAGAAACAGCCTGGCTCCCCATCCCCTCCCAGTGTTCCAGAGTCGGGACA 540
G S P E K Q P G S P S P P S V P E S G Q
GGGTGTAACCAAGGGTGAAGGGGGAACTTCAGTCCCAGCTTCCCTGCCTGGGGGTAGCAA 600
G V T K G E G G T S V P A S L P G G S K
GGAAGAGGAGGAGAAGGCTAAGCGTCTGCTCTACTGTGCACTGTGCAAGGTGGCTGTGAA 660
E E E E K A K R L L Y C A L C K V A V N
CTCCCTGTCCCAGCTTGAGGCACATAACAAAGGTACTAAGCACAAGACAATTTTGGAGGC 720
S L S Q L E A H N K G T K H K T I L E A
CGGAAGTGGGCTGGGAGCCATCAAAGCTTACCCTCGGGTGGGGCCTCCAATTCTGGGGAA 780
G S G L G A I K A Y P R V G P P I L G N
CCAGAGGCTCCTGCCCAGGGACGGAACCTTCCACTGTGAGATCTGCAATGTCAAGGTCAA 840
Q R L L P R D G T F H C E I C N V K V N
TTCGGAGGTCCAGCTGAAACAGCACATCTCCAGCAGGAGGCACCGAGATGGCGTGGCTGG 900
S E V Q L K Q H I S S R R H R D G V A G
GAAGCCCAACCCTCTACTGAGCCGGCACAAGAAGCCTAGGGGCGCTGCAGAGCTGGCGGG 960
K P N P L L S R H K K P R G A A E L A G
CACGCTGACTTTCTCAAAGGAGCTGCCCAAGTCCCTGGCCGGTGGCCTGCTCCCCAGCCC 1020
T L T F S K E L P K S L A G G L L P S P
CCTAGCGGTGGCTGCGGTGATGGCCGCTGCAGCAGGATCTCCGCTGTCCCTGCGTCCAGC 1080
L A V A A V M A A A A A G S P L S L R P A
TCCAGCTGCACCTCTTCTGCAGGGACCACCGATCACACACCCTCTACTCCACCCTGCCCC 1140
P A A P L L Q G P P I T H P L L H P A P
AGGACCCATCCGAACTGCGCACGGACCCATCCTCTTCTCCCCCTACTGACCTCAACCCTG 1200
G P I R T A H G P I L F S P Y *
AACCCCTCCCATTGAATCCCCCACCTCCAGCCGGGACCCAGGTCTCCGGGCTCCCAGCCC 1260
GCCCCTCTTCCCGGCGCTCCCTGAATGGTCTCTCCTTCCCCCCACCCCGAGGTACGGGGT 1320
TCCAGGAAAGGGGAGGGGTAGCGGGGGAGGGGGGCTTCAGAAGGGGGGGAACACCCCAGA 1380
TCTCAGGAACCCCGCCCCCTGCCCTTCCCTCTCCCCTAGAAAAGAGGGGGGCCGTCTCAC 1440
CCCCGAGCCCCCTCGGAGACACCCCCTCCCAAAAGCCATGTCCATCCAACCCTTCCCCTC 1500
CAAACCTAGCACAAAACGGGGTTCACAAGCCATGGTCGGGGCCCGGGGGGAGAACATGGA 1560
TTTTCTTGGCAATAAGCGGACTCTGGGACTCCGGCTCCCCTACCCCAAACTAAAGCGCTT 1620
CCGTGAACAACCCCATCCTCCCGCAGGAGGAGGGGAACAGGCGGGATCCTGGGTCCCTCG 1680
TAAGCACTTTGGTTTTACCGCCTGCAACCTCACTGTGCCCGCCCCGCACCATGCCCCAGC 1740
CCCGGGTCCAGTCGGGCCCATCGCAGGGGGCAGCGCTTGGGGGCATCTCAGGCACTTGGG 1800
TGGGACCAAGGAGATGCCCTCATAGACCCTTCCCTCGCCTTCTTCCTCCCCGGTCCGGGT 1860
TCCATTCTTTTCACCAGCACCCATCGCCCAAGGGGTACCAAGGGGGCAAGGGGTGTCCA 1920
GTCCAAGCCCACCCCCGCCTCGCCTTCCGCAAAACTGTGAGCAAAAAGCAATAGAAGCCT 1980
CGCCCCCGCCCCCGGCCCGTCCCTTTCCGCAGGATTTGCCGTCTGTAGCCTCCCCATTCC 2040
AGTTCCTAGACCTCATGGCTGCCCCCTCCCGTGACACCTCCACTGCACAACTCGGGCGGG 2100
GGCGTGACCCCCAACCCGTCCCTTCTGTGGTTTCCGTGTGGTCAGCCCTTCCAGCCGCCG 2160
ACCCGGGATGGGGACTGGCCCCCACTGGCCCTCCCCTCCCCATGGACTCTTTCTGCTTGA 2220
CAATGTAGCAAACCCCCTCCCTCTTTCTCTCTCTTTGACAATAAAGTCTGGATTTGTTCT 2280
GCCCTCTCGCCAAGGAAAAA 2300

FIGURE 7

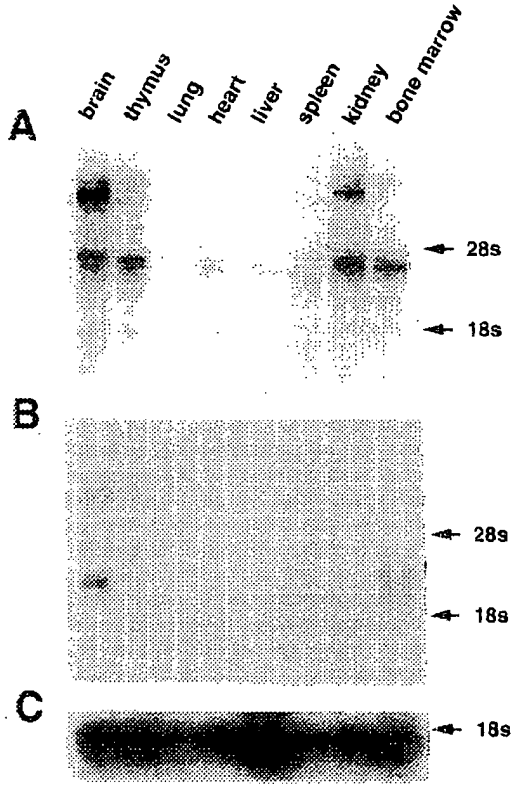
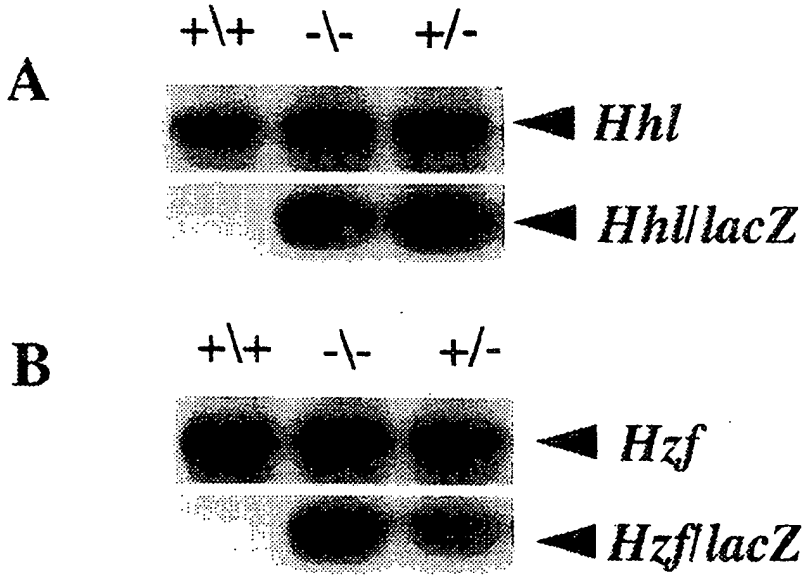

brain
thymus
lung
heart
liver
spleen
kidney
bone marrow
A
28s
18s
B
28s
18s
C
18s

FIGURE 8



SEQUENCE LISTING

<110> Mount Sinai Hospital

<120> NOVEL HEMATOPOIETIC GENES AND POLYPEPTIDES

<130> P170PCT9

<140> not yet issued

<141> 2000-02-18

<150> US 60/120,972

<151> 1999-02-19

<160> 14

<170> PatentIn Ver. 2.0

<210> 1

<211> 2297

<212> DNA

<213> Murinae gen. sp.

<400> 1

cggaaccccg ggagccaggg tccccagcat gatcctcggc agcctgagcc gggcggggcc 60

cctccctctg ctccggcagc cccccatcat gcagccaccg atggacctca agcagatcct 120

gcccttccca ctagagccag ccccaaccct gggcctcttc agcaactaca gcacaatgga 180

ccctgtacag aaagctgtgc tctcccacac ttttggagga cccttgctca agaccaagcg 240

gccagtcatt tcctgtaatg tctgtcagat ccgcttcaat tctcagagcc aggctgaggc 300

gcactacaag ggtaatcgcc atgcccgaag agtcaaaggc atcgaagctg ccaaaacccg 360

aggcagggag cctagtgtcc gggaatcagg agatccagct ccagcaggca gcatccctcc 420

gagtggggat ggtgtagccc ctcgtccagt ttccatggag aatggcctgg gtccagctcc 480

aggatcccca gagaaacagc ctggctcccc atcccctccc agtgttccag agtcgggaca 540

gggtgtaacc aagggtgaag ggggaacttc agtcccagct tccctgcctg ggggtagcaa 600

ggaagaggag gagaaggcta agcgtctgct ctactgtgca ctgtgcaagg tggctgtgaa 660

ctccctgtcc cagcttgagg cacataacaa aggtactaag cacaagacaa ttttggaggc 720

cggaagtggg ctgggagcca tcaaagctta ccctcgggtg gggccaattc tggggaacca 780

gaggctcctg cccagggacg gaaccttcca ctgtgagatc tgcaatgtca aggtcaattc 840

ggaggtccag ctgaaacagc acatctccag caggaggcac cgagatggcg tggctgggaa 900

gcccaaccct ctactgagcc ggcacaagaa gcctaggggc gctgcagagc tggcgggcac 960

gctgactttc tcaaaggagc tgcccaagtc cctggccggt ggcctgctcc ccagcccccc 1020

agcggtggct gcggtgatgg ccgctgcagc aggatctccg ctgtccctgc gtccagctcc 1080

agctgcacct cttctgcagg gaccaccgat cacacaccct ctactccacc ctgccccagg 1140

acccatccga actgcgcacg gacccatcct cttctccccc tactgacctc aaccctgaac 1200

ccctcccatt gaatcccccа cctccagccg ggacccaggt ctccgggctc ccagcccgcc 1260

cctcttcccg gcgctccctg aatggtctct ccttcccccc accccgaggt acggggttcc 1320

aggaaagggg aggggtagcg ggggaggggg gcttcagaag ggggggaaca ccccagatct 1380

caggaacccc gccccctgcc cttccctctc ccctagaaaa gaggggggcc gtctcacccc 1440

cgagcccccc cggagacacc ccctcccaaa agccatgtcc atccaaccct tcccctccaa 1500

acctagcaca aaacggggtt cacaagccat ggtcggggcc cgggggggaga acatggattt 1560

tcttggcaat aagcggactc tgggactccg gctcccctac cccaaactaa agcgcttccg 1620

tgaacaaccc catcctcccg caggaggagg ggaacaggcg ggatcctggg tccctcgtaa 1680

gcactttggt tttaccgcct gcaacctcac tgtgcccgcc ccgcaccatg ccccagcccc 1740

gggtccagtc gggcccatcg caggggggcag cgcttggggg catctcaggc acttgggtgg 1800

gaccaaggag atgccctcat agacccttcc ctcgccttct tcctccccgg tccgggttcc 1860

attcttttca ccagcaccca tcgcccaagg ggtaccaagg ggggcaaggg gtgtccagtc 1920

caagcccacc cccgcctcgc cttccgcaaa actgtgagca aaaagcaata gaagcctcgc 1980

ccccgccccc ggcccgtccc tttccgcagg atttgccgtc tgtagcctcc ccattccagt 2040

tcctagacct catggctgcc ccctcccgtg acacctccac tgcacaactc gggcgggggc 2100

gtgaccccca acccgtccct tctgtggttt ccgtgtggtc agcccttcca gccgccgacc 2160

cgggatgggg actggccccc actggccctc ccctccccat ggactctttc tgcttgacaa 2220

tgtagcaaac cccctccctc tttctctctc tttgacaata aagtctggat ttgttctgcc 2280

ctctcgccaa ggaaaaa 2297

<210> 2

<211> 385

<212> PRT

<213> Murinae gen. sp.

<400> 2

Met Ile Leu Gly Ser Leu Ser Arg Ala Gly Pro Leu Pro Leu Leu Arg
1 5 10 15

Gln Pro Pro Ile Met Gln Pro Pro Met Asp Leu Lys Gln Ile Leu Pro
20 25 30

Phe Pro Leu Glu Pro Ala Pro Thr Leu Gly Leu Phe Ser Asn Tyr Ser
35 40 45

Thr Met Asp Pro Val Gln Lys Ala Val Leu Ser His Thr Phe Gly Gly
50 55 60

Pro Leu Leu Lys Thr Lys Arg Pro Val Ile Ser Cys Asn Val Cys Gln
65 70 75 80

Ile Arg Phe Asn Ser Gln Ser Gln Ala Glu Ala His Tyr Lys Gly Asn

85 90 95

Arg His Ala Arg Arg Val Lys Gly Ile Glu Ala Ala Lys Thr Arg Gly
100 105 110

Arg Glu Pro Ser Val Arg Glu Ser Gly Asp Pro Ala Pro Ala Gly Ser
115 120 125

Ile Pro Pro Ser Gly Asp Gly Val Ala Pro Arg Pro Val Ser Met Glu
130 135 140

Asn Gly Leu Gly Pro Ala Pro Gly Ser Pro Glu Lys Gln Pro Gly Ser
145 150 155 160

Pro Ser Pro Pro Ser Val Pro Glu Ser Gly Gln Gly Val Thr Lys Gly
165 170 175

Glu Gly Gly Thr Ser Val Pro Ala Ser Leu Pro Gly Gly Ser Lys Glu
180 185 190

Glu Glu Glu Lys Ala Lys Arg Leu Leu Tyr Cys Ala Leu Cys Lys Val
195 200 205

Ala Val Asn Ser Leu Ser Gln Leu Glu Ala His Asn Lys Gly Thr Lys
210 215 220

His Lys Thr Ile Leu Glu Ala Gly Ser Gly Leu Gly Ala Ile Lys Ala
225 230 235 240

```
Tyr Pro Arg Val Gly Pro Ile Leu Gly Asn Gln Arg Leu Leu Pro Arg
                245                 250                 255

Asp Gly Thr Phe His Cys Glu Ile Cys Asn Val Lys Val Asn Ser Glu
            260                 265                 270

Val Gln Leu Lys Gln His Ile Ser Ser Arg Arg His Arg Asp Gly Val
        275                 280                 285

Ala Gly Lys Pro Asn Pro Leu Leu Ser Arg His Lys Lys Pro Arg Gly
    290                 295                 300

Ala Ala Glu Leu Ala Gly Thr Leu Thr Phe Ser Lys Glu Leu Pro Lys
305                 310                 315                 320

Ser Leu Ala Gly Gly Leu Leu Pro Ser Pro Leu Ala Val Ala Ala Val
                325                 330                 335

Met Ala Ala Ala Ala Gly Ser Pro Leu Ser Leu Arg Pro Ala Pro Ala
            340                 345                 350
Ala Pro Leu Leu Gln Gly Pro Pro Ile Thr His Pro Leu Leu His Pro
        355                 360                 365

Ala Pro Gly Pro Ile Arg Thr Ala His Gly Pro Ile Leu Phe Ser Pro
    370                 375                 380

Tyr
385
```

<210> 3

<211> 1258

<212> DNA

<213> Murinae gen. sp.

<400> 3

ctcgggcgtc tgacagagtt gttgttccgg cacgccgcgg agacgtgagt tgttccaggt 60

ggatgtgggc agagaggttt gcagaactga aatggaggtc ggagcttcgt tccagaaggt 120

tagtgggtca tctgattctg tggccacact gaacagtgaa gaatttgttt tggtttctca 180

gcacacagat gccacttcta taaaggatga tgggaagcca cagctgaaga tagcttccaa 240

tggtgatgag cagttggaaa aagccatgga agagattttg agagattccg agaaaggaca 300

aagcggtcta cctgttgatt gccaaggatc cagtgagatt tcagactgtc cttttggaga 360

tgtgccggcc agccaaacaa ctaagccgcc tctccagtta attttggatc catctaatac 420

agaaatttcc acacccagac catcttctcc aagcagattt cctgaagaag acagtgttct 480

ctttaacaag ctgacatact taggatgtat gaaggtttct tcccacgcag tgaagtggag 540

gctttacggc catgccacca tgagagcttc cagtcagtac ccctttgctg ttactctgta 600

tgtgcccaat gttccagaag gatctgtgag aatcatagac cagtcaagca atgtggagat 660

agcatctttt ccaatttata aagtgctttt ctgtgcacgt gggcatgatg agacagccga 720

gagcaattgc tttgcattta cagagagttc tcatggctca gaagaatttc agatacatgt 780

tttctcctgt gaaattaaag aggcagtaag cagaatttta tatagtttct gcactgcatt 840

caaacgttct tccagacaag tgtctgatgt taaagactca gtcatcccga ctcccgacag 900

tgatgtgttt accttcagtg tctccttgga ggtcaaagaa gatgatggaa aaggaaactt 960

tagtcccgtg cctaaggata gagataaatt ttatatcaaa ataaagcaag gaatagagaa 1020

gaaggttgtg attacagttc agcaactgtc taacaaagaa ttagctattg agagatgttt 1080

tggaatgtta ttaagcccag gtcgaaacgt gaagaacagt gacatgcatt tactggatat 1140

ggagtcccgg aaagagctat gatgggagag cttacgtcat cacgggcatg tggaacccca 1200

acgcaccaat atttctggct cttaatgaag aaaccctgtc tcaaaaaaaa ccaaaaaa 1258

<210> 4

<211> 298

<212> PRT

<213> Murinae gen. sp.

<400> 4

```
Met Glu Glu Ile Leu Arg Asp Ser Glu Lys Gly Gln Ser Gly Leu Pro
  1               5                  10                  15

Val Asp Cys Gln Gly Ser Ser Glu Ile Ser Asp Cys Pro Phe Gly Asp
             20                  25                  30

Val Pro Ala Ser Gln Thr Thr Lys Pro Pro Leu Gln Leu Ile Leu Asp
         35                  40                  45

Pro Ser Asn Thr Glu Ile Ser Thr Pro Arg Pro Ser Ser Pro Ser Arg
     50                  55                  60

Phe Pro Glu Glu Asp Ser Val Leu Phe Asn Lys Leu Thr Tyr Leu Gly
 65                  70                  75                  80

Cys Met Lys Val Ser Ser His Ala Val Lys Trp Arg Leu Tyr Gly His
                 85                  90                  95

Ala Thr Met Arg Ala Ser Ser Gln Tyr Pro Phe Ala Val Thr Leu Tyr
            100                 105                 110

Val Pro Asn Val Pro Glu Gly Ser Val Arg Ile Ile Asp Gln Ser Ser
        115                 120                 125
```

```
Asn Val Glu Ile Ala Ser Phe Pro Ile Tyr Lys Val Leu Phe Cys Ala
    130                 135                 140

Arg Gly His Asp Glu Thr Ala Glu Ser Asn Cys Phe Ala Phe Thr Glu
145                 150                 155                 160

Ser Ser His Gly Ser Glu Glu Phe Gln Ile His Val Phe Ser Cys Glu
                165                 170                 175

Ile Lys Glu Ala Val Ser Arg Ile Leu Tyr Ser Phe Cys Thr Ala Phe
            180                 185                 190

Lys Arg Ser Ser Arg Gln Val Ser Asp Val Lys Asp Ser Val Ile Pro
        195                 200                 205

Thr Pro Asp Ser Asp Val Phe Thr Phe Ser Val Ser Leu Glu Val Lys
    210                 215                 220

Glu Asp Asp Gly Lys Gly Asn Phe Ser Pro Val Pro Lys Asp Arg Asp
225                 230                 235                 240

Lys Phe Tyr Ile Lys Ile Lys Gln Gly Ile Glu Lys Lys Val Val Ile
                245                 250                 255

Thr Val Gln Gln Leu Ser Asn Lys Glu Leu Ala Ile Glu Arg Cys Phe
            260                 265                 270

Gly Met Leu Leu Ser Pro Gly Arg Asn Val Lys Asn Ser Asp Met His
        275                 280                 285
```

Leu Leu Asp Met Glu Ser Arg Lys Glu Leu
290 295

<210> 5
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:primer

<400> 5
gcagagctct gggcggcggg t 21

<210> 6
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:primer

<400> 6
ctcgggcgtc tgacagagtt g 21

<210> 7
<211> 18
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 7

cccgggacag aacgacgc 18

<210> 8

<211> 18

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 8

cggaccccgg gagccagg 18

<210> 9

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 9

attacgccag ctggcgaaag g 21

<210> 10

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 10

gattggtggc gacgactcct g 21

<210> 11

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 11

tatcttcagc tgtggcttcc c 21

<210> 12

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 12

ccgtcgactc tggcgccgct gctctgtcag 30

<210> 13

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 13

cggaaccccg ggagccagg 19

<210> 14

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:seq14

<400> 14

tctggggatc ctggagctgg ac 22